Annals of Theoretical Psychology 12

Craig W. Gruber
Matthew G. Clark
Sven Hroar Klempe
Jaan Valsiner *Editors*

Constraints of Agency

Explorations of Theory in Everyday Life

Springer

Constraints of Agency

Annals of Theoretical Psychology

The *Annals of Theoretical Psychology* is devoted to understanding theoretical developments and advances in psychological theory. This series is designed to further the dialogue on theoretical issues in the field of psychology and to unify the discipline through a theoretical synthesis of ideas on key issues of debate. Core themes of the Annals vary from one volume to another, moving beyond a focus on one particular aspect or approach to theory. Each book consists of invited and submitted papers and commentaries that explore a facet of innovative theory in psychology. Of particular interest is moving the discussion and exploration of theory into application for use in research, practice and teaching, taking into account the globalized nature of contemporary psychology.

The enduring objective of the Annals of Theoretical Psychology is the exploration of key concepts that require further inquiry, dialogue, and theoretical integration within psychology and related fields.

For further volumes:
http://www.springer.com/series/5627

Craig W. Gruber • Matthew G. Clark
Sven Hroar Klempe • Jaan Valsiner
Editors

Constraints of Agency

Explorations of Theory in Everyday Life

Springer

Editors
Craig W. Gruber
Department of Psychology
Northeastern University
Boston
Massachusetts
USA

Sven Hroar Klempe
Department of Psychology
NTNU
Trondheim
Norway

Matthew G. Clark
United States Military Academy
Westpoint
New York
USA

Jaan Valsiner
Niels Bohr Professor of Cultural Psychology
Department of Communication and
Psychology
Aalborg University
Aalborg
Denmark

ISBN 978-3-319-10129-3 ISBN 978-3-319-10130-9 (eBook)
DOI 10.1007/978-3-319-10130-9
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014950146

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Introduction

While sitting on a train with my friend and colleague Hroar Klempe, we began discussing the role of agency and agentic behavior in an individual. Our discussion included items such as: what is the nature of the Agency, the neuroscience behind it, and how does psychological thought inform and be informed by agentic processes. As the train picked up speed along the tracks, we came to the conclusion that true agency comes from constraints that are put on agency. In the absence of constraints, individuals can become overwhelmed, with truly unlimited options from which they can choose. As the train shifted tracks and shunted along the rails, our conversation continued and ranged on to assess why this was so.

This seeming incongruous conclusion led to a discussion of the differing ways in which that might be felt or experienced by people. As we continued along the rails, we began to explore these situations and circumstances. As an historian of psychology, Hroar began asking the questions and exploring through a myriad of thought experiments that how agency was and might have been explored and discussed in psychology's evolution. While he and I have some quite definitive thoughts on the matter, we believe that Roger Smith has addressed them from a truly unique perspective and are quite happy about the ensuing dialogue.

Subsequently, we also discussed the role of neuroscience in agentic behavior. Our discussion led further to conversations which I had with Matthew Clark. He and I have spent countless hours discussing, arguing, and then agreeing to explore further the nature of the neuroscience of agency. In our discussions, Matt took the lead in addressing this area of inquiry. While reviewing our discussions (which were both passionate and good natured), I came to see that the viewpoints and topics we discussed are quite well presented in the section led by William Klemm. I believe that this section allows us to look into the neuroscience of agency in a way in which we have not looked, as a discipline, before. It is exciting for me, as a psychologist and one of the editors of the *Annals* project, to see this discussion taking place among academics of varying views.

The editorial help by Lauren Takakjian and Dakota Snyder for all of this volume is acknowledged with great gratitude.

Lastly, as the train continued toward the end of the line, Hroar and I took our discussion to the role of agency in the psychology of religious thought. It was a fascinating conversation in which we discussed the role of psychology in religion, regardless of one's religiosity. The concept of agency in religion is one which has been intriguing me for quite some time. Free will, agency, psychology, are all terms which have varying levels of acceptance and applicability to individuals depending on their faith background, or a lack of faith background at all. To lead the discussion, Phil Helsel brings an opening point to the discourse which I believe is both thought provoking and discerning. The commentaries that follow, I believe, help to take the conversation of agency in religion to a place from which further research, writing, and discourse may easily flow.

As our train reached its terminus, Hroar and I ended our discussion of the moment, promising to continue it via dialogue and writing. We enlisted the help of Matt Clark and Jaan Valsiner to make it happen, and the four of us present it to you here.

I believe it is a testament to surround ourselves by those who will inform our conversation, provoke our thoughts, and stretch our cognitive skills that we have assembled as the authors and contributors you see in this 12th edition of the *Annals of Theoretical Psychology*. As we wrote in Volume 11, we strongly believe that the place for discussion and theory in psychology is here, and the time for these discussions is now, as we strive to build and recognize theoretical psychology as a discipline which does not negate the theories which have brought psychology to where it is today, but rather to rely upon the theories and theoreticians who have gone ahead of us to develop the next generation of psychologists. This next generation will only be able to move forward by standing on the shoulders of the psychologists and theoreticians who have come before us to lay the groundwork for new thoughts, as they did for the generation of psychologists before them.

Please, join the conversation.

June 2014 Craig Gruber

Contents

Contributors

Sarah H. Awad Aalborg University, Aalborg, Denmark

Robert Childs Massachusetts School of Professional Psychology, Newton, MA, USA

Matthew G. Clark United States Military Academy, West Point, NY, USA

Vlad Petre Glăveanu Aalborg Universitet, Aalborg, Denmark

Craig W Gruber Northeastern University, Boston, MA

Philip Browning Helsel Boston College, Chestnut Hill, MA, USA

William R. Klemm Texas A&M University, College Station, TX, USA

Sven Hroar Klempe Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Markus E. Schlosser University College Dublin, Dublin 4, Ireland

Roger Smith Moscow, Russia

Luca Tateo Niels Bohr Centre for Cultural Psychology, Aalborg University, Aalborg, Denmark

Frederick Toates Open University, Buckinghamshire, UK

Jana Uher Department of Psychology, London School of Economics and Political Science, London, UK

Comparative Differential and Personality Psychology, Free University Berlin, Berlin, Deutschland

Jaan Valsiner Niels Bohr Professor of Cultural Psychology, Department of Communication and Psychology, Aalborg University, Aalborg, Denmark

Brady Wagoner Aalborg University, Aalborg, Denmark

Jing Zhu Sun Yat-sen University, Guangzhou, China

About the Editors

Craig Gruber is a social and developmental psychologist who is a lecturer at Northeastern University. He earned his B.A. from American University, an M.S. from Johns Hopkins University and his Ph.D. from Clark University. His research and theoretical interests reside in the area of courage, personal agency, and educational settings. Another primary area of interest for Craig is his interest in teaching psychology, especially, in developing curricula and the application of psychological theory in the classroom. In addition, his writing interests involve psychology's history as a resource of ideas for contemporary advancement of the theory and teaching. He is the co-author of Psychology, AP* Edition (Pearson) with Phillip Zimbardo and founding editor of the Whitman Journal of Psychology, a high school journal of student research and writing. He has also papers published in Integrative Psychological and Behavioral Science, Culture and Psychology and Teaching of Psychology.

Matthew G. Clark is a behavioral neuroscientist at the Department of Behavioral Sciences and Leadership at the United States Military Academy at West Point, New York. He received a B.A. in Psychology from Coe College, Cedar Rapids, Iowa, and a Ph.D. in Behavioral and Neural Sciences from Rutgers University in 2000. As a behavioral neuroscientist he conducted research on developmental language disorders, learning, motivation, psychopharmacology, and medical countermeasures against chemical warfare nerve agents. In 2002, his research was recognized as the top research at the Singapore International Symposium on protection against toxic substances (SISPAT). He worked on urgent medical and military science and technology issues at various locations around the world, including in a theater of combat. He has published six book chapters across these various areas, including contributions to the "Textbook of Military Medicine, Medical Aspects of Chemical Warfare", "The 71F Advantage: Applying Army Research Psychology for Health and Performance Gains", and "Pharmacological Perspectives of Toxic Chemicals and Their Antidotes." He also has over 25 referred publications and professional reports.

Sven Hroar Klempe is the associate Professor in Psychology at the Department of Psychology, NTNU, Trondheim, but he has also a background as full Professor

in musicology, Asst. Prof. in media studies, teacher, journalist and music critic. His educational background is diverse and primarily from Norway, and also from France. His research is cross-disciplinary with an emphasis on theory of science, communication, and the history of psychology.

Jaan Valsiner is the Niels Bohr Professor of Cultural Psychology at Aalborg University in Denmark. He is the founding editor (1995) of the Sage Journal of Culture & Psychology and Editor-in-Chief of Integrative Psychological and Behavioral Sciences (Springer, 2007). In 1995 he was awarded the Alexander von Humboldt Prize for his interdisciplinary work on human development.

About the Authors

Sarah H. Awad is a Ph.D. student at the Centre for Cultural Psychology, Aalborg University. She received her M.Sc. in social and cultural psychology from London School of Economics and Political Science and her B.A. in mass communication from the American University in Cairo. Her research interests are in the interrelation between the fields of cultural psychology, communication, and social development. She studies the process by which individuals develop through times of life ruptures and how art is used as a symbolic resource in shaping their agency through the experience.

Robert Childs is in the teaching faculty of the Massachusetts School of Professional Psychology where he teaches courses on spirituality and Jungian psychology, humanistic and existential psychology and on clinical supervision. Having worked in clinical practice in the Boston area for many years, Dr. Childs's clinical specialties include working in the psychology of adoption, working with artists from various mediums on creativity and psychological process, and working with the psychology of twinship. He received his Psy.D. from the Massachusetts School of Professional Psychology, a Master's degree from the Harvard Graduate School of Education in counseling psychology and was a fellow for two years at the Harvard Divinity School.

Vlad Petre Glăveanun is the Associate Professor at the Niels Bohr Professorship Centre for Cultural Psychology, and Co-Director of Center for Creativity Research, Aalborg University. He received his Ph.D. from London School of Economics.

Philip Browning Helsel works in the field of pastoral theology, a discipline that gained prominence in the 1940s and 1950s at the intersection of psychology, theology, and the social gospel movement in mainline American seminaries. Through the influence of clinical pastoral education—a supervised psychological training program for clergy in hospitals—pastoral care and counseling became a core part of the theological curriculum. Pastoral care and counseling is frequently taught at seminaries and divinity schools in hopes of preparing persons to become more thoughtful, socially aware, and healthier in a variety of church and social service settings. His current work focuses on how the power of psychiatric diagnosis shapes pastoral care and counseling.

Bill Klemm began his career as a physiologist, interested in the brain. As a charter member of the Society for Neuroscience, he quickly adopted the interdisciplinary modality of neuroscience, and has in recent years tried to integrate a variety of disciplines into his thinking about agency. He has published provocative papers on such topics as nerve impulse codes, EEG signatures of thought, sleep and dreaming, mechanisms of behavior, and free will. He co-authored Brainstem Mechanisms of Behavior (Wiley), and his latest thoughts are captured in the newly released book, Mental Biology: The New Science of How the Brain and Mind Relate, distributed by Random House.

Roger Smith is an emeritus Reader in History of Science at Lancaster University, UK, and an independent scholar living in Moscow, Russia, where he is an associate of the Institute of Philosophy and the Institute of Psychology of the Russian Academy of Sciences. He is a past President of the European Society for the History of the Human Sciences and a founding Associate Editor of the journal, History of the Human Sciences.

Markus E. Schlosser is a Lecturer in Philosophy at University College Dublin. He has worked and published extensively on various topics in the philosophy of action, the philosophy of mind, and the philosophy of neuroscience (especially the neuroscience of free will).

Frederick Toates is an emeritus Professor of biological psychology at the Open University in the UK, where he researches the control of behavior by means of applying systems principles. He has held visiting professor status in France, Germany and Sweden. He has theorized extensively on incentive motivation and hierarchies in the control of behavior and has a particular interest in the role of dopamine in motivation. His latest books are Burrhus F. Skinner (2009), Biological Psychology (2011) and How Sexual Desire Works: The Enigmatic Urge (2014).

Luca Tateo is a member of the Niels Bohr Professorship Centre on Cultural Psychology at Aalborg University, Denmark, where he leads the core on epistemology and methodology. He is a Marie Curie Pst-doctoral Fellow and has for long time also worked at University of Salerno in Italy.

Jana Uher is a Research Fellow at the London School of Economics and Political Science. She received her Ph.D. from Free University Berlin in 2009, where she set up and headed the research group Comparative Differential and Personality Psychology from 2010–2013, funded by the Deutsche Forschungsgemeinschaft (DFG). In 2013, Jana Uher was awarded a Marie Curie Fellowship by the European Commission. Her research is transdisciplinary, concentrating on philosophy-of-science issues of psychological and behavioral research on individuals from culture- and species-comprehensive perspectives. To demonstrate her metatheoretical and methodological developments, she has been working empirically with human children and adults with different sociocultural backgrounds and more than ten different nonhuman species including the great apes, capuchin monkeys and macaques. Jana Uher employs a broad portfolio of methods including noninvasive experimental

studies, behavioural observations, and audiovisual and computerized methods for contextualized investigations. She has also been conducting studies with adults and human observers, respectively, about human impression formation using standardized assessments of both human and nonhuman individuals. Her empirical research projects are rooted in international and national collaborations with scholars from different disciplines. She has been working at the Max Planck Institute for Evolutionary Anthropology in Leipzig (2003–2005), and has also been a visiting scholar at the Institute of Cognitive Sciences and Technologies, National Research Council of Italy (ISTC-CNR) in Rome (2011–2013).

Brady Wagoner is Professor with special responsibilities at the Niels Bohr Professorship Centre for Cultural Psychology, Aalborg University. He received his Ph.D. from the University of Cambridge, where he was a Gates Cambridge Scholar and co-creator of the F.C. Bartlett Internet Archive and journal Psychology & Society. He is also associate editor of culture & psychology and peace & conflict. His books include Symbolic Transformation (Routledge, 2010), Dialogicality in Focus (Nova, 2011), Culture and Social Change (Info Age, 2012), Development as a Social Process (Routledge, 2013) and Cultural Psychology and Its Future (Info Age, 2014). He is currently working on a single authored book titled The Constructive Mind: Frederic Bartlett's Psychology in Reconstruction (Cambridge University Press) and editing the Oxford Handbook of Culture and Memory.

Jing Zhu is Professor in the Institute of Logic and Cognition and Department of Philosophy, Sun Yat-sen University. His research areas include philosophy of science, philosophy of mind and action, epistemology, and moral psychology. He has published in Philosophical Studies, Synthese, Canadian Journal of Philosophy, Journal of Consciousness Studies, Consciousness and Cognition, and Philosophical Psychology.

Part I
Historical backgrounds on agency

Chapter 1
Agency: A Historical Perspective

Roger Smith

1.1 Delineating Agency

'Agency' is a word with multiple meanings. As it certainly does not ambiguously denote a psychological category, I begin with clarifications.

The word 'agent' in English has been in use since the seventeenth century to identify a factor or power held to cause a change. For 'agency', the Oxford dictionary cites Darwin, who wrote about the pollination of flowers 'requiring the agency of certain insects to bring pollen from one flower to another'. 'Agency' denotes capacity and power attributed to matter (as in chemical 'reagent'), to institutions or social organizations with the power to act on behalf of people (like 'the Central Intelligence Agency'), and to people individually (as in 'the agent of her own destruction'). It has also been common to refer to God's agency or the agency of spirits. Historically, the opposite of 'agent' was 'patient', and this is a reminder of the active/passive distinction of the premodern ontology to which the language of agency is heir.

The multiplicity of references to agents of different kinds persists in everyday speech about causation. The language of agency permits descriptions of naturalistic and non-naturalistic, material and mental, individual and social causes to exist alongside and in interaction with each other. In everyday usage, it is possible to refer to everything from car brakes to rental company to driver error to intention to kill as the principal agent of an accident. However, there has been a historical shift, so that there is a modern emphasis on using the word 'agency' to denote a moral or political, that is to say, distinctively human category. As a result, speakers now might reserve the attribution of agency to people or institutions in a car accident (to continue with the example) and distinguish other factors as contributing causes. This modern usage describes the agents as individual people or groups of people who are said to have the power to be the cause of events, and said to have the power of self-direction. Agency has become linked to notions of the autonomous self and

R. Smith (✉)
Obolenskii per. 2-66, Moscow 119021, Russia
e-mail: rogersmith1945@gmail.com

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_1

to the dignity or status accorded to a 'free agent'. Thus, it is a notion important to moral and political issues. Feminists, for example, place great weight on women acquiring agency and in critiquing the circumstances in which this is constrained. Indeed, much of the experienced meaning of agency derives from its opposition to the notion of constraint. Beyond this, the notion of agency also has an influential place in humanist and existential philosophy, with expression in psychology, where reference to agency denotes something like a reference to freedom as a defining condition of being human.

It is important to be aware of the range of linguistic usage. It is possible to refer both to material or spiritual things as agents and to people, distinctively, as agents. Before the twentieth century, language sometimes described people as agents, but in doing so it ordinarily attributed the agency to the will, to the soul, or to reason, rather than to a psychological subject or a self. In a parallel way, some recent writers attribute agency to the body, to the unconscious, or to the brain. All the same, contemporary writers commonly attribute agency to people, or to the self (as in the statement, 'I did this'), and they attach a special value to it. These multiple usages are rather confusing for analysis, if unexceptional in everyday speech.

In the light of these comments, I stress one point. Reference to agency in twentieth or twenty-first century psychology, as in the human sciences generally, may simultaneously invoke what are generally thought of as causal processes and what are thought of as free actions. At first glance, as a result, it would seem as if psychologists are deeply equivocal about accepting or denying free will. I suggest that actually there is no deep-lying confusion behind this equivocation, if that it be; rather, there is something special in the projects of psychology, namely, their ability to provide description and analysis appropriate *for understanding people,* as opposed to understanding brains (in terms of deterministic causal processes) on the one hand, and juridical, moral, or political subjects (with imputed absolute freedom of action) on the other hand.

The modern notion of the *individual person as agent* first developed in legal, political and theological contexts. The history is intimately connected to the developing notion of a self. Reference to human agency denoted, and still denotes, action originated by individual legal, moral, and political subjects, or by institutions viewed as analogous to individual subjects, acting without special or noteworthy constraint. This has often been called free agency. The Oxford dictionary, in this context, cites Coleridge's political demand that 'the State shall leave the largest portion of personal free agency to each of its citizens, that is compatible with the free agency of all'. It is the normative practices of politics and morality that have made it important to distinguish human agents from other agents. Insofar as psychologists have been drawn into discussions of human agency, they have taken part in these normative practices.

In consequence, it makes no sense to ask for a non-evaluative account of a person's agency. Discussion of agency, in the last analysis, involves questions about the relation of human subjects, intentions, and evaluations, and of the language and culture which are their expression, to the causal material world that is the subject

of the natural sciences. This is why agency is a problematic category for natural scientists who think that science excludes evaluative judgments.

When psychologists refer to agency, they use a psychosocial category. Just as there is no non-evaluative use of the word 'agency', the word has no psychological meaning independent of social content. The literature about agency, as a psychosocial category, therefore inevitably takes positions on the long-standing question of the relation between psychological and social forms of explanation and the institutionalization of those forms in separate psychological science and social sciences. The discipline of social psychology clearly faces these matters most directly (and there has even been a psychological social psychology and a sociological social psychology; Good 2000).[1]

While ordinary people, and often enough psychologists too, might now say that a person *is* an agent (and explain the agency, say, by reference to intelligence or cognitive capacities), the social scientist understands agency as *attributed to* a person (or institutions or things). In the language of social science, agency is a *status* not a *state* (Barnes 2000).[2] In a related way, in the language of many analytic philosophers, will, agency, and choice are not powers but human actions under certain kinds of descriptions. From the viewpoint of these disciplines, it is a misusage, a category mistake, to talk as if agency were a psychological state or psychological power. Rather, it is a power of persons (the classic statement is Ryle 1949/1963). Applying this lesson in psychology, we can say that when psychologists talk about agency, they utilize a category with social content and take part in the process of social and political ascription of status to people. But it muddling matter. This is because reference to agency as a status ascription (or attribution) persists alongside and interacts with the older usage in which reference to an agent denoted a capacity or power (whether material, mental or spiritual). Moreover, there are, of course, psychologists with religious beliefs or who uphold a humanist philosophical anthropology, for whom agency is indeed a 'real' state, a state valued and thought essential to being fully human.

There is nothing contradictory, then, though it may be confusing, to describe the *body* as the agent of a person's desires, while at the same time describing something in the body (an illness perhaps) as constraining a *person's* agency. By virtue

[1] For an overview of the relations of the varieties of psychology to natural science, Smith, 2013b.

[2] Barry Barnes, a social theorist, makes the case for a naturalistic understanding of agency as assigned status. He is sceptical of the psychologist's practice of attributing agency to 'internal' mental, or cognitive states on the one hand, and he is critical, on the other hand, of the way the group of social scientists known as ethnomethodologists distance theory from material practice and in effect implement a dualism separating research on humans from research on nature. Another, quite different but influential, approach to understanding the attribution of causes is the actor–network theory initiated by Michel Callon and Bruno Latour in the sociology of science. This theory, which aims to understand why any particular piece of knowledge acquires authority, treats all relevant factors, human and non-human alike, as actors (or, we might say, agents) in the negotiation of knowledge claims. For Latour (2005; Law & Hassard, 1999), an 'actor' is simply what makes a difference. Actor–network language goes against the grain of contemporary usage of the concept of agency in normative statements about 'the human', but, interestingly, is compatible with early modern and still common usage, identifying agency with cause.

of its flexibility and open-endedness, everyday language about agency has a rich instrumentality.[3] For example, it makes sense to talk about training or disciplining the body in order to build up patterns of behaviour (or habits) understood in causal terms, in order to give a person more agency, understood as freedom of action. Contemporary advocates of the brain sciences promise that new knowledge (of causal processes) will give individuals more agency (freedom of action; Rose 2007, 2013). Indeed, this is the common pattern of argument of the psychological and social sciences over the past couple of centuries: Let us understand human nature (causally) in order to improve human wellbeing (free agency). All the same, there have been and there are religious and humanist critics of this, the enlightenment project, who hold, for instance, that the very act of investing in causal explanatory language about people derogates from the morality and politics of free human agency.

To the extent that agency, understood as a category that is properly applied to persons, is now a flourishing interest for psychologists, this, surely, is yet another manifestation of the individualism characteristic of Western modernity. Ways of life informed by psychology are bound up with concern for the power and capacity individuals have and do not have in the social worlds they inhabit. It might be thought that as researchers in a field of natural science, psychologists could and should substitute the notion of cause for the notion of agency in describing and explaining behaviour. As a family of fields of research about people, and even more as a family of practices concerned with everyday individual capacity and activity, however, psychology has a large place for the category of agency. Even if psychologists have at times carried on as if the human subject were not inherently social (as if brains existed in vats, for example, not in social people), this is not possible for those who adopt the language of agency. It is most obvious, perhaps, in psychotherapy—in all its multifarious forms. Therapeutic reference to agency invokes knowledge of the 'internal' powers and constraints of a person, along with the 'external' powers and constraints of the social world the person inhabits.

Notions of agency and constraint are of manifest significance in everyday psychology, as a conversation taking place between two of the people involved with the volume in which this chapter appears illustrates. (Whether my account is accurate is not important now, and I have used my imagination.) Their interest was whether lack of constraint on individual actions in at least some areas of Western society, allowing individuals the freedom 'to do their own thing', or, as the advert says, 'just do it', might not actually be a constraint. Intense individualism is indeed not easy for many people. The discussion was not about either autonomous will or neurons but enlarged on a psychological, everyday approach to the *psychosocial* conditions of people's lives. It was talk about people, not wills or brains, as agents. Such talk has a history and a social specificity.

The complexity of language about agency is compounded when there is an implied reference (as in the citation from Coleridge) to *free* action. Indeed, I rather

[3] My argument here is admittedly impressionistic. I find support in both discursive psychology and ethnomethodology, research areas that illuminate a seemingly endless flexibility of world construction in which assigning causes (and agency) plays a large part.

think connotations of free action always colour the word 'agency'. I do not propose to tie myself or my readers up in the free will/determinism debate. But, I will note how psychological as well as everyday language about agency often combines description of a person as the site of power bringing about an effect (analogous to saying water is the agent of dissolution of a salt) and the description of a person as a *free* agent (for example, a person responsible in law). The language of agency does not dictate a position on the free will/determinism question: it holds things open and allows further evidence and judgment to refine description and explanation. Everyday usage allows for fudge, continuous negotiation, infinite shades, in attributing agency understood as both cause in the scientific sense and as the kind of free power people attribute to people able to carry out legal and moral acts. There are, to be sure, radical libertarians who emphasise the extent of agency understood as individual free power. And there are severely reductionist scientists, by contrast, for whom there can be no such thing as free agency. Many people in Western cultures, I presume, take a position somewhere between these extremes, and in doing so, they think of human agency as the activity of a person (relatively) without hindrance or constraint from other powers, a person who has experience of variable degrees of actual agency. In this middle position, it is common for 'agency' to mean something like 'the real cause' or 'the principal cause' of something happening, that is, the cause which on a particular social occasion is thought to be the key to the actual unfolding of events. Thus, someone might refer to the brakes not the driver as the agent, the real cause, of an accident, though of course both brakes and driver belong in a full list of causes. (The full list, literally speaking, requires the history of the world.) Philosophers discuss what I am calling the broad middle position under the heading of 'compatibilism', the label for the view that it is not inconsistent to accept determinist causal explanations in the natural (or social) sciences while upholding belief about the free action of people (Kane 2005).

I am suggesting that there are constructive *psychological* ways of referring to the agency of individuals, individuals with the power or capacity thought intrinsic to being human in general and to being the persons they are in particular. This psychological usage contrasts with polarized alternatives. These alternatives represent the two poles of Cartesian dualism: attributing agency to mechanistic body, and hence attributing all change to the physical agency of matter in motion; or attributing agency to a soul imbued with transcendent powers. Psychological reference to agency, at its best, displaces this polarity. Ordinary speech in psychological society, and much psychological discourse both lay and professional, refers actions *to people* not to the operation of mechanistic processes or souls. It uses language incompatible, for instance, with the belief, upheld in some versions of popular neuroscience, that there is no human agency or personal volition because 'we are our brains'. But psychological reference to agency does not presuppose that agency is the expression of a specific power of the soul or faculty of mind, like the will, understood as somehow acting independently of physical processes. The psychological approach to agency is historically an outgrowth from, and continues to blend with, everyday understandings of the powers that people have and the constraints under which these powers operate. These understandings have multiple roots in political,

religious, ethical, legal and philosophical culture. Only in recent centuries, perhaps we might say in the past two centuries, have these roots nourished a distinctively psychological representation of the agency.

Modern studies of agency as an attribution recognize its intrinsically social content, and, with that, recognize the intimate connection of attribution of agency and ascription of *responsibility*. In many contexts, certainly in judicial settings, to attribute agency *is* to ascribe responsibility. As the British sociologist Barry Barnes argues: 'An understanding of the everyday employment of this concept [of responsibility], with its double significance —psychologically understood it implies internal capacities, sociologically understood it implies liability and answerability—is…the key to an understanding of the role of 'choice', 'agency' and related concepts in everyday contexts' (Barnes 2000, p. 1). Consider also the word 'Victorian'. It is a word now coloured by knowledge of ardent nineteenth century convictions about moral agency, which was thought to be exercised through individual strength of will, strength of character, and by the fact that the presence or absence of will was thought to validate judgments about the responsibility of individuals. It is worth bearing in mind the insightful, if exaggerated, comment of one late Victorian writer, who said that it is *only* the social question of responsibility and punishment that sustains public interest in the question of free will (Hyslop 1894, pp. 181–182; Smith 2013a, p. 165). Very often, we will find that when the talk is about agency, the talk is at base about responsibility.

1.2 Volition and Psychological Agency

There are many ways of conceptualizing and representing agency. Earlier ages treated spiritual powers, the devil, the stars, and the passions as agents. There are, of course, modern communities that continue to make such attributions. The history of specifically psychological conceptions of agency encompasses the field of volition, desire and will—the division of conation, alongside cognition and affection, in traditional tripartite description of mind. In this section, I contribute a historical sketch.

There is a curious interest in this for the modern psychologist because volition largely disappeared from view as a topic for professional psychologists during much of the twentieth century. Since the 1970s, there has been some revival (as called for in Kimble & Perlmuter 1970). Experimental psychologists transferred an interest in agency, understood as the motor of behaviour, from mental will to drive, personality, cognitive processes, the emotions, the unconscious, or, alternatively, to the body—and in recent decades to the brain. Large areas of psychology went ahead without concern for or even with active antagonism to the notion of volition.[4] Yet at the same time, therapeutic, educational, organizational, and coun-

[4] A point made in Daston, 1978, in comments on US experimental psychology, in relation to which she framed views of the emergence of psychology in late nineteenth-century Britain. I discuss the

selling psychological practices all deployed some notion of psychological agency, even of volition, even if implicitly rather than in an openly systematic and theorized form. Notions of volitional agency flourished even more in the domain of so-called popular psychology, in the world of know yourself and self-help books. In the contemporary world, and above all in the world of imputed consumer choice ('just do it') there is a strong focus on agency, agency which in ordinary speech often enough appears as an expression of will. The word 'agency' is indeed widely used to denote individual power in actual or desired forms of economic exchange and governance. Most significantly, in respect of social power, legal systems continue to pose decisions about responsibility in terms which people understand as referring to internal mental states of volitional agency. Presumably, the renewed interest among some psychologists in volition as a category reflects all this.

The early intellectual history centres on belief in the intrinsically active power of the soul to cause actions. Christians in the early church sought knowledge of the will in the light of their understanding of the relationship between the will of God (with omnipotent agency) and the action of the human soul (owing to sin, with restricted agency). The question of freedom of the will, that is, the extent of human agency independent of God's grace, was a source of deep divisions from the time of the later Stoics and of Augustine (Frede 2011). It came to a head again during the Reformation, and it was the subject of a famous exchange between Erasmus and Luther.[5] There was a large Renaissance literature on more secular views of volition, particularly as applied to the control of the passions. Shakespeare's plays commented on the role of fate or *fortunà*, as opposed to the will, in a person's life. Belief in the power or agency of the individual soul was a source of dignity and, for some, the very basis of civilized, as opposed to barbarian, existence. In his discussion of the passions, Descartes wrote: 'I see only one thing in us which could give us good reason for esteeming ourselves, namely, the exercise of our free will and the control which we have over our volitions. For we can reasonably be praised or blamed only for actions that depend on this free will' (Descartes 1649/1985, p. 384).[6] There was

late nineteenth-century British debate on volition in detail in Smith, 2013a, where I suggest that different conceptions of what a science of psychology should be were at stake rather than opposition between science and unscientific moral conceptions of volition. That volition might feature and then disappear from view as an acknowledged psychological category is suggestive for the history of psychological categories generally. There is a historiography on whether such categories as memory, intelligence and emotion, as well as volition, and even the category psychology itself, should be taken to be 'timeless', or whether they have a history (as I certainly think). The issues are very complex and I must leave them aside now. But see Danziger, 2008; Smith, 2005.

[5] The exchanges between Erasmus and Luther are usefully brought together in Erasmus–Luther 1961/2007. For a broad-ranging discussion (though corrected historically in Frede, 2011), Arendt, 1978.

[6] Heidegger's (1987/2001, pp. 117–19) account, in seminars with therapists, held that it was Descartes' idea of the ego which led to the metaphysics of the subject/object distinction and hence to the whole modern problem of locating subjectivity in relation to causal events.

an association between the will, agency, dignity and emerging notions of the identity and worth of the individual self.[7]

The actual word 'agent' spread with the expansion and increased precision of natural philosophy in the seventeenth century. The word denoted powers or capacities without drawing sharp distinctions between spiritual, mental or material agency. Old theological disputes and new science together ensured the continuation of debate about the place of freedom and 'necessity' in the course of things. The word 'determinism', applied to nature and to human affairs, came into use only after about 1870 (Hacking 1983).

Locke's prime concern in his *Essay Concerning Human Understanding* (1690) was the sources and certainty of knowledge. His book nevertheless included a very influential statement on the origin (or agent) of human action, which he attributed to pleasure and pain. Empiricist writers subsequently pictured individual agency as the response of a person to the pleasures and pains of experience.[8] Whether, and in what sense, this implicated necessity was much debated. Some authors, like Joseph Priestley in England and C.-A. Helvétius in France, embraced necessity and, indeed, they built their hopes for enhanced individual political and moral agency, and thus for enlightened political and social life on this basis. Writing in the same spirit, Jeremy Bentham established political utilitarianism, a scheme for a legal system which would appropriately distribute pleasure and pain, and hence regulate individual agency and thereby ensure a rational ordering of human affairs. He expected people to be rational and hence to choose his system: he assumed that natural human rationality conferred agency. The many opponents of utilitarianism argued, however, that belief in necessity was destructive of morality and incompatible with common-sense awareness of each individual's power to will actions. They also doubted the extent to which rationality in fact conferred moral or political agency. The more conservative of these critics turned to the moral will, informed by religious faith and enforced by royal and religious power, as the source of agency. Indeed, in the eighteenth and nineteenth centuries, there was an evangelically informed turn in Britain and in the USA to the authority of subjective knowledge of agency, understood as a mental power, 'the will'. Most authors in English, at least before the second third of the nineteenth century, did not conceive of these discussions as being about psychological agency in any distinguishable sense; rather, their subject was human nature in all its social, political, psychological, economic, legal, medical, linguistic, and philosophical dimensions. Agency was distributed as a subject in what was called moral philosophy (and, later, the moral sciences), and there was no specific discussion of agency as a category in its own right.

[7] An alternative approach to the history of agency might begin with the notion of the self, in order to tie the history directly to social, legal, economic and political thought (Seigel, 2005; Taylor, 1989). It seems to me that attempts to write the history of the self spin off into an uncontainable range of topics. The same would be the case were anyone to undertake anything so rash as 'a history of agency'.

[8] I draw here and in what follows on the broad history in Smith, 1997. The more specifically psychological dimensions of the story are rewritten and updated in Smith, 2013b.

A profound and, in the long run, influential expression of belief in the intrinsic agentive character of the mental world, which later commentators did not hesitate to call a contribution to psychology, appeared in the writings of Maine de Biran in the first decade of the nineteenth century. Biran judged awareness of effort to be the source of the most elementary and irreducible knowledge of mind, and he therefore rendered mental agency—'l'effort voulu'—constitutive of the personal mind or self.[9] He published piecemeal and he left his essays unfinished; nevertheless, later writers understood his work to originate a distinctively French psychological–philosophical view of agency, developed most prominently in Bergson's and in Sartre's (completely different) accounts of free agency as defining what it was to be human.

The work of Victor Cousin was actually more influential in France in the first half of the nineteenth century. It acquired formal standing in the French higher education system as a result of Cousin's position at the head of the institutional structure for teaching teachers. Making a broad claim for 'psychologie' as the road to philosophy, Cousin, like Biran, stressed the irreducible volitional character of the self; but whereas Biran used language expressive of an almost phenomenological awareness or subjectivity, Cousin used a more recognizably traditional Christian language referring to the power of the soul. When Taine and Ribot turned against this psychology in the 1870s, promoting what Ribot called 'the new psychology' (an eclectic mixture of German experimental psychology, physiological psychology, clinical evidence, and British associationism), the stage was then set, as elsewhere at the time, for debate about the implications of developments in mental science for agency and determinism.[10]

According to a least one historian of philosophy, Kant bequeathed the antimony of freedom and necessity as '*the* great problem of modern thought' (Pinkard 2002, p. 43). When he discussed the nature of human beings, Kant separated 'the human' as an anthropological subject, for which empirical argument was appropriate, and discussion of the essentially human as the agent of the moral law, for which he turned to transcendental reason. His anthropological writings and lectures (which subsumed psychology) discussed human nature and activity in a manner that flowed together with everyday talk in the educated circles in which he and his students moved. He included discussion of capacity and agency in connection with habit, mental disorders, constraint, character, and so forth. His formal moral discourse, by contrast, defined the condition of being human in terms of an abstract, absolute imperative, and in this context, freedom denoted the human obligation and power to use reason and act according to the moral law. This propounded a moral theory of obligation and not a psychological theory of capacity or agency. All the same, Kant's arguments underwrote and legitimated an understanding of agency as a capacity of the human spirit, the position that informed German-language

[9] An accessible account is in Biran, 2005. I provide sources for a history of the sense of effort and movement in Smith, 2011.

[10] In general, Carroy, Ohayon, & Plas, 2006. Goldstein, 1994, related Cousin's psychology to the question of the place of agency in Michel Foucault's understanding of history—which some accused of leaving no place for agency.

philosophical anthropology in the nineteenth century and beyond. The claim that agency, embedded in the human spirit and expressed through language and culture, was constitutive of being human lasted till, but hardly survived, the horrors of the twentieth century. It certainly fostered psychological formulations, as one can see, for example, in C. G. Jung's project in scientific psychology (in which Jung sought to juxtapose his own conclusions about the collective unconscious with Kant's thought; Shamdasani 2003).

When John Stuart Mill published *A System of Logic,* the book that provided the philosophical underpinnings for utilitarian social and political thought, he dealt directly with the issue of necessity. He took the position that was to be the mainstay of modern philosophical argument (the position, as noted above, called compatibilism). In Mill's presentation of the case, a person's actions were said to be free when they were the actions of the whole person, unconstrained in any significant way. Certainly, Mill held, pain or pleasure was the proximate cause of a movement or of behaviour, but this fact was compatible with saying that a *person* had agency (or, as the Victorians would have said, showed character and exercised will) if the movement or conduct expressed the purposes and character of the whole person (Mill 1843/1900, Book VI, Chap. 2). A person in prison did not have agency in the way a person at home did, even if both responded to pleasures and pains. Mill was a political libertarian and a moralist committed to creating conditions that would enhance personal agency, but he believed in both necessity and personal freedom while strongly opposing any notion of the will as some kind of spiritual force.

During Mill's lifetime, that is, through the middle years of the nineteenth century, a large specialist and popular medical and scientific literature spread the conviction that a person's agency might not be nearly as extensive as those who stressed the role of the will, Mill included, tended to assume. There was considerable interest in instincts, habits, and automatism, in the hypnotic state and in spiritualism, and in disorders implying a loss of control from drunkenness to epilepsy. A physiological psychology developed, and its promoters built on the model of reflex action to propose new scientific understanding in bodily terms of much that had earlier been thought attributable to the activity of mind. In this connection, the London physiologist W. B. Carpenter, in 1853, introduced the idea of what he called ideomotor action, to describe the way an idea in the mind, or an obsessive thought, caused activity over which a person had no control and of which a person may even have been unconscious (Carpenter 1853, p. 672). In effect, Carpenter (and other medical writers) redescribed the mental act of anticipation leading to a movement as a physical (higher brain) reflex. The physiological language merged with a large popular literature, replete with vivid case studies, that discussed when and where individual agency, understood to depend on will, as an empirical matter of fact did or did not exist (Smith 2013a). There was an English language automatism debate, prompted by T. H. Huxley's lecture on the topic in 1874, and this was the intellectual background of William James's intense personal and theoretical psychological interest in the will (Huxley 1874/1894). James took over Carpenter's language of ideomotor action and argued that action follows directly from the presence of an idea in the mind and does not require a separate mental effort or will force (James 1890/1950,

vol. 2, pp. 522–528). He set this psychophysiological discussion of will in a larger framework of argument, however, and this larger framework advanced into another version of compatibilism. While stressing the physical determinants of psychological life (as he did in his theory of the emotions as well as of the will), James upheld 'the will to believe' as a moral project, indeed as a project without which he as a person could not live well, and this emphatically defended a notion of individual agency.

Though it was a cliché of the time to describe the nineteenth century as an age of science, belief in free will did not just persist in this age but even gained in strength from what was thought to be empirical knowledge. A large part of the popular appeal of phrenology in the 1820s and 1830s, for instance, was that people believed knowledge of cerebral capacities as the determinants of character and gave individuals power to strengthen or control mental life for themselves (Wyhe 2004). Though accused of materialism and hence immorality, phrenology increased not decreased a sense of personal agency for those who thought it true. Analogously, in the course of the evolutionary debates, leading advocates of scientific naturalism like Huxley and Francis Galton promoted belief in the uniformity of causation in nature and human nature as part of a moral crusade to empower individuals and society alike on the basis of 'the facts'. It was the message of Huxley's much commented on lecture, 'Evolution and Ethics', which ended by quoting lines from the late English national poet, Tennyson: 'We are grown men, and must play the man "strong in will/ To strive, to seek, to find, and not to yield"' (Huxley 1893/1989, p. 86). Science supported a moral project of enhancing agency. Galton's studies of the contribution of heredity to character, which had a determinist form, were matched by a political programme, eugenics, to do something about it. Medical demonstrations of automatism in illnesses like epilepsy, in hypnotic trance, or in habitual drunkenness, fed into Victorian rhetoric about the necessity of will power—the cultivation of personal agency.

There were numerous philosophically oriented attempts to overcome what seemed to many people to be confusions of thought in this mixture of scientific and moral culture. This debate about the nature and possibility of free will in a scientific age was important to the way psychology developed as a field in Britain, Europe, and the USA . It sustained links between psychology and mental philosophy just when some psychologists were looking towards experimentation as the means to make psychology a distinct scientific field. An integrated commitment to scientific knowledge *and* willed agency was a striking feature of the late Victorian age.

If we turn to the everyday language of human relations, the language, for example, of the nineteenth-century novel, we find a mixture of the attribution to causes (habit, social conditions, sex, heredity, and so on) and attribution to individual will to be not just ordinary but ubiquitous. From the 1830s or so, English-language writers increasingly, but never exclusively, articulated what they had to say about such topics under the heading of psychology. Victorian authors continuously mixed and negotiated description of mental forces, such as will power, and causal determinants, such as training, age, social circumstance, illness, and custom. Stirring the mixture often enough supplied the novelist with a plot. It was a mixture of

seemingly infinite variety. While holding that young children were not agents and hence not legally responsible, both Catholics and Evangelical Protestants stressed the presence of the will in children and held it to be an innate moral or spiritual force, a force for both good (as self-help texts presupposed) and bad (as references to wilfulness and to upbringing designed to break the child's will made clear).[11] A large moralistic literature about individual agency attended to the practical powers that individuals might possess, acquire, or lose. Strictly consistent explanation in the language of mental forces or in the language of causal bodily processes was not a priority. Discussion of agency was marked by flexibility of description and openness to negotiation according to particular circumstances that observers thought prevailed. There were, to be sure, times where the flexibility broke down, and I discuss in the next section the criminal court where a defence of insanity sometimes congealed and polarized opinion.

There was a noteworthy struggle over the representation of agency in Tsarist Russia, which influenced the way psychology developed in that country (Sirotkina & Smith 2012). The rigidly autocratic political system upheld faith in the soul as the bearer of personal agency and ground for attributing individual responsibility. Social order in the autocracy appeared to require this faith. During the brief period of relative liberalization under Alexander II, in 1863, I. M. Sechenov published the first version of his article on 'The Reflexes of the Brain' and N. G. Chernyshevsky published a notorious novel, *What Is To Be Done?* Sechenov, a physiologist trained in German and Austrian laboratories, turned to the model of reflex action in order to imagine a physiological analogue for the mental process of volition. When criticized for removing the grounds for belief in individual responsibility, he denied that this was the implication of his argument. For Sechenov, knowledge of the natural bodily conditions of life was the basis for the exercise of agency, not faith in the soul. He did not use a political language of agency or even a moral language of free will; indeed, he was not permitted to do so under conditions of censorship.[12] All the same, he contributed to a political programme to replace the theocratic agency of the one tsar with the enlightened rational agency of many informed subjects. Even though his language described nervous and psychological processes, readers understood the message. Chernyshevsky's novel was more direct and, remarkably, still published (Chernyschevsky 1863/1989). His story brought to life a group of young people who had adopted a full-fledged rational egoism, the position that for Chernyshevsky expressed reasoned agency. He attributed to reason the agency to act on behalf of the natural needs of the person, to unite body and mind, and to escape the fetters of irrational passion and ignorant and repressive moral and religious codes. Liberated individual actions informed by reason, he maintained, would ensure collective progress.

[11] For the spread of popular psychological practices, Thomson, 2006; and on the psychology of children, Shuttleworth, 2010.

[12] For his own account of this, Sechenov, 1965. A version of 'Reflexes of the Brain' first appeared in English (Sechenov, 1935/1968) in 1935, and there was a French version in 1884; it was not known to non-Russian speakers when first published.

These writings were part of a public debate, in the course of which they faced two of the most widely read and influential ripostes to such notions of agency in the nineteenth century. Interestingly, in everyday language now it would be a commonplace to say that these ripostes showed much greater psychological insight, even if their authors did not in any way describe themselves as psychologists or contribute to academic psychology. Turgenev, in his novel *Fathers and Children*, the publication of which preceded Sechenov's and Chernyshevsky's contributions, portrayed a student of medical physiology, Bazarov, who unsentimentally dismissed the possibility of any kind of agency that did not have the form of a physical agent (Turgenev 1862/1960). His own agency as a proponent of this materialist worldview, however, was tragically cut short when he fell in love and when he contracted an infection that killed him. His own conception of agency was pathetically inadequate for his own life. Independently, Dostoevsky, enraged by Chernyshevsky's novel, in a response poorly understood at the time because it appeared so antihumanist, wrote *Letters from the Underworld*. The *Letters* purported to be written by a man (significantly with no name) who, in opposition to any conceivable constraint, and most painfully in opposition to the gift to him of actual, profound love, asserted the arbitrariness of his own will (Dostoevsky 1865/1960; Frank 1986, Chap. 21). Dostoevsky's antihero was an agent, come what may. The *Letters* thus bitterly parodied rational egoism—they pictured the will as self-destructive as well as destructive of others. Dostoevsky's view of agency was radically anti-enlightenment and pictured agency as a source of tragedy in the human condition.

Within the family of Aleksandr I. Herzen, the Russian political exile, there was actually a living and not fictional debate between father and son. The father, who belonged to the romantic generation coming to maturity in the 1830s, faced by necessitarian physiological argument, turned to human history and social progress to legitimate a conception of agency. His son, Alexandre A. Herzen, flushed with enthusiasm for the physiological science of the 1850s (and a future professor of physiology in Lausanne), upheld determinism (Sirotkina 2002).[13] Neither was a psychologist in any specialist sense; but such debate about agency over the years played a large part in shaping psychology as a public field.

When experimental psychology began to acquire a place in German universities, followed by other countries, its academic proponents inherited the antimony of freedom and necessity. Wundt, to take a key instance, maintained a separation between causal analysis thought appropriate at the level of psychophysiological phenomena, and a form of understanding in terms of mental apperceptive and conative activity though appropriate at the level of higher mental events (Danziger 2001). He elaborated a psychology with, in effect, a place for active, agentive mind or spirit, and this psychology existed alongside the contributions he made to physiological psychology. He developed research in the former under the heading of *Völkerpsychologie*,

[13] Interestingly, the son was in sympathetic communication with the most forthright English exponent of physiological determinism, the specialist in mental disorders, Henry Maudsley, who was also a fierce moralist (further illustrating the everyday mixture of reference to agency and to causes in speech).

a science that turned to language, myth, and cultural life generally as the collective expression of human agency. In other hands, especially at Würzburg under the leadership of Külpe, by contrast, in the years before 1914, there was research to make volition a rigorously examined experimental topic, though one could hardly describe this as in any direct way about agency. Albert Michotte also studied in detail the immediate antecedents of voluntary choice (studies which preceded his better known work on the perception of causation; Michotte & Prüm 1911). With the establishment of a psychology discipline in universities and colleges in the USA, which involved considerable concern for the scientific standing of the field, volition gradually became sidelined as a topic. North-American psychologists turned to the motivation, intelligence or personality, rather than volitional agency, of their subjects, to represent activity in ways judged to be scientific.

For the German idealists, Fichte, Schelling, Hegel, and (for these purposes) Schopenhauer, and in the books of many more accessible exponents of idealist worldviews, like E. von Hartmann, agency was not a specifically psychological topic. Rather, idealists sought to characterize the dynamic form of existence as such and only secondarily to explain the agency of a particular person as an individual. Idealist thought, however, underwrote belief that there *is* agency, though this was variously understood to have the form of impersonal powers (perhaps unconscious forces, perhaps a 'will to power'), or to have the form of a psychological expression of spirit in the individual mind. Awareness of the potential conflict between agency understood as reason and agency understood as non-rational force contributed to the literature of intellectual crisis, which Nietzsche did so much to deepen, that was so prominent around 1900.

Elements of the idealist intellectual tradition persisted and influenced humanistic forms of psychology in the post-1945 period. Existentialist thought stressed ontological freedom, the ultimate agentive character of the human condition and this was taken up in psychological terms, for example, in the work of Erich Fromm. Informed by Christian ethics rather than European philosophy, Rogerian therapy, built on the client-centred principle of nondirective regard, was a practical enactment of the agentive ontological status attributed to clients.

What was experienced as an intellectual crisis, because the foundations of reason themselves appeared to be questioned, was the context of new thought, in great variety, about the unconscious. Freudian psychoanalysis strongly emphasized the agency of the unconscious, rather than the agency of rational conscious capacities, in action and character formation. In contrast to earlier notions of unconscious events, which pointed to the role of unconscious anticipation, the new psychologists of the unconscious pointed to the burden of repressed memory (individual and inherited; Hayward 2014). Freud and his followers, indeed, portrayed this as a new step in human self-understanding. It is therefore very striking that Freud, who was committed to a strong psychological determinism (complementing physiological determinism), also remained committed, like a good Kantian philosopher, to the power of reason to stand apart from unconscious forces, to comprehend them, and thereby to offer at least some hope for human freedom (Tauber 2010, 2013). Freud acquired a reputation as the hammer of bourgeois confidence in the power of the

will, since people are, he claimed, driven by unconscious forces; yet, at the same time, he reasserted a Kantian view of agency as the possession of reason to discern the moral law. In this way, I would argue, he exemplified enlightenment thought as it developed in the psychological and social sciences. In these sciences, a notion of agency persisted, the agency of the reasoning subject given institutional form in the world of science, even while the activity of scientific reason was creating knowledge exposing the causal determination underlying what people did. The resulting dilemmas have been acknowledged and explored in political thought. For example, Sonia Kruks has persuasively discussed Simone de Beauvoir's work as a long engagement with the political ambiguity of the human subject understood as *both* causally situated and agentive, *both* conditioned and free (Kruks 2012, Chap. 1). De Beauvoir's art in exploring this as an irreducible ambiguity, finding a literary as well as philosophical voice, brought what she had to say close to the practical understanding of people in everyday relations. This ambiguity persists, I would affirm, in the modernist project, which asserts the causal necessity of events alongside the demand that individuals live one way rather than another.

The philosophical conundrum here has given rise to a huge literature, much of it refining different stances in the position known as compatibilism.[14] Ordinary people continue by and large to be compatibilists, as Barnes notes: 'Much of our everyday discourse manifests a robust compatibilism, in that it is content to regard actions as at once chosen and caused' (Barnes 2000, p. 4).

Modern psychological discussion of free agency, in line with compatibilist argument, has taken the form of empirical studies of everyday ways of ordering the world *in terms of perceived causes*. Social psychology research in the second half of the twentieth century shifted towards the study of the cognitive and emotional processes involved in people's judgment and understanding of the agentive character of what people did. This research therefore did not directly address the question—which came to appear ontological rather than scientific in nature—as to whether people *were* agents. Debate was not about what agency 'was' or to what extent people were 'really' free agents, but about how people viewed agency in terms of personal traits (including their own) varying between individuals and even varying within one person in different circumstances and over time. Attribution theorists, for example, in the 1970s and 1980s, researched what people perceived and said in everyday life in order to understand causal cognition and attendant moral judgment (Jaspars, Fincham, and Hewstone 1983; Hewstone 1989).[15] It would seem that this brought social psychology somewhat closer to a social science approach, in which agency was understood as a status attribution. There was, besides, interest in finding

[14] I try here to keep clear of philosophical discussion of the question of free will. It is a labyrinth in which analytic philosophers have staked out a multitude of highly refined positions. If agency is discussed as assigned status, the philosophical issues are not pressing.

[15] The founding of attribution theory as a social psychology was the work of Kelley, 1967, 1971, building on Heider, 1944.

psychological tools to increase personal agency. Beyond this, there was discussion of the personal and social advantages of belief in agency.[16]

The dismissal by twentieth-century psychologists of volition as a category was part of a general suspicion about explaining human capacities by internal mental states or processes. Many sociologists, often influenced by Wittgenstein, shared this suspicion and understood references to mental states as a certain kind of language game or social activity. The sceptical psychologists supposed that human activity was at base the activity of physical systems: there was no mental power, certainly no free will.[17] The sceptical sociologists argued that references to mental states were status attributions, descriptions of social relations in particular ways of life. From the sociological point of view, agency was a collective achievement expressed in social order built on the regulatory notion of individual responsibility. For large numbers of psychologists, agency was a function of physiological capacities. For one large group, the social psychologists whom I have just mentioned, research was interested in what people believed rather than in agency as a 'real' state. Yet, in spite of all this and whatever the power of the arguments, everyday psychological discourse, and the discourse of a good deal of expert psychology along with it (therapy, guidance, forms of training, and such like) continued in fact to refer, routinely, to internal states and, if not so much specifically to will or volition, to desire, intention, motive, purpose, and choice.

Both history and ethnology provide a comparative perspective on when and how psychological practices advance or damage agency—and of *which kinds of agency and for whom*.[18] The work of the sociologist Nikolas Rose is an influential reference point in discussion of liberal democracies (Rose 1985, 1989/1999, 1996/1998). Under the rubric of 'the history of the present', Rose described psychological practices as distinctively modern forms of governance in societies of a kind in which many people, but very far from all, have come to live, that located the power or agency maintaining social order *within* individuals and *within* occupations (psychotherapy, educational guidance, counselling, etc.) that worked to ensure people were responsible individuals. During the course of the twentieth century, the Victorian emphasis

[16] Here, in part, I paraphrase comments made about the paper in an anonymous reader's report. But it is for other people to describe directions in current research.

[17] In the last two decades, argument has acquired an empirical dimension based on the experiments of Benjamin Libet, experiments that, on some interpretations, show awareness of free will to be 'an illusion' (Wegner, 2002; along with critical comment in Pockett, Banks, & Gallagher, 2006; McClure, 2012; Rodder & Meynen, 2013; Tallis, 2011, pp. 51–59, 247–256). The large literature this work has generated has rather disguised, it seems to me, the important point that will is something which people in *some societies* attribute to persons on *particular social occasions*. Such attribution is not a matter for scientific psychology but a matter of the right use of language in appropriate social settings (Bennett & Hacker, 2003, pp. 224–231).

[18] In much discussion of agency and free agency, there is a painful silence about the wealth of evidence that different societies categorize and classify actions (and even the human and animal boundary) in markedly different ways. It is simply wrong to imply that 'all people' have this or that particular view of agency. (E.g. introduction to Baer, Kaufman, & Baumeister, 2008, p. 3: 'In general, however, people implicitly assign a sense of agency and of free will, to themselves and others.') For informative exploration of the riches and complexities of taking a comparative perspective, Lloyd, 2007, 2009.

on will and will power was reshaped as advice and training in the techniques of self-management. When such psychological practices were at work in governance on a broad front, we can identify psychological society, a society in which there is a strong tendency to assign agency to individuals understood as psychological beings (Smith 2013b, Chap. 4). Rose was substantially concerned with the question where agency was ascribed—conscious that this linked the history of psychology to politics. His arguments, sociological in form, located agency not in individual selves or minds but in the practices themselves. 'To account for the capacity to act one needs no theory of the subject prior to and resistant to that which would capture it—such capacities for action emerge out of the specific regimes and technologies that machinate humans in diverse ways' (Rose 1996/1998, pp. 186–187).[19] There have been humanist critics who saw such analysis (which was linked both by Rose and by his critics with the work of Michel Foucault) as damaging notions of individual agency; but Rose argued that agency lay with the enhancement and availability of practices not with some imagined 'internal' reality (Rose 2007; Derkson 2011).[20]

The history of psychology is also a source of studies of the agency of individual subjects of psychological practices—the patients in Freud's case studies, for instance, or the participants in psychological experiments, some famous, like 'Little Albert' (Harris 1979, 2011) or like the students in Milgram's research on obedience.[21] These studies open up the politically and ethically significant topic of how in fact psychologists have themselves ascribed agency to the people with whom they interact, with what consequences, and what co-operation or opposition from participating subjects (or 'patients'). This contributes to a critical, politically and ethically reflexive, psychology.

1.3 The Insanity Defence: Debate on Agency Exemplified?

In this section, I try briefly to provide a more precise historical case study of what debate about agency may mean in practice. Exemplary demonstration of a number of the points made in this essay is to be found, perhaps, in an area where statements

[19] 'Machinate', I think, is jargon for the way the body becomes part of instrumental systems.

[20] It was an element of Derksen's response to Gergen, 2010, to point out the instrumentality of psychological practices (e.g. discipline, meditation, positive thinking) in increasing the range of a person's choice or agency—agency is, and is well known to be, *variable*. Foucault discussed power as the activity of everyday practices and of ordinary bodies, rather than as a function of the top-down organization of society, and it would seem to have been his view that we could identify agency in the life of these everyday practices—putting the lie to the view that he allowed no space for it.

[21] Historical work on Freud's case studies has proved to be an extremely critical tool in showing the mythology in Freud's construction of psychoanalysis. See Borch-Jacobsen and Shamdasani, 2012, Chap. 3. Historical work on 'Little Albert' and the obedience experiments has shown the importance of the history of the institutionalization of ethical practices in the psychology profession.

with psychological (or psychiatric) content interact with legal decision making. I say, 'perhaps', because the question of agency in connection with the plea of insanity in criminal cases (the focus of the debate here) has attracted a huge amount of comment, and comment has returned again and again to the issues as if the ground had not been gone over many times before. Clearly, the issues continued to be troubling.

Western codified and common law criminal legal systems maintain that the establishment of guilt requires a demonstration both that a certain thing happened (*actus reus*) and that the accused party was an actor in the events with a certain capacity of mind (*mens rea*).[22] In everyday language, we might say that a finding of guilt requires a defendant to have been the agent of the relevant event.[23] As a result, over the centuries, there has been debate about what makes a person a *legal agent,* and statements have centred on mental states and capacities. Legal writers have construed these states and capacities variously, sometimes as formal legal categories with no specific empirical referent, sometimes as empirically verifiable states of mind, intentionality, or social attribution. They have understood the discussion to be a matter for jurisprudence, not at base, a matter for psychology.

When physiological approaches to mind became common in the nineteenth century, creating psychological medicine, proponents sometimes said that science had disproved the existence of free will (in individual cases or even in general) and, as a result, the administration of justice and punishment had to change. The same argument is occasionally heard now in the wake of the huge growth of research in the neurosciences. The argument was, and is, vitiated by at least two misapprehensions. First, statements that causes determine effects are not empirical conclusions of natural science but logical expressions of the form of knowledge established in natural science. All events, scientifically understood, have causes: finding the causes is the name of the game. This holds just as much for sociological or psychological as for physiological research. Second, it follows, free will is not a meaningful notion in natural science discourse (or in social science or psychological discourse understood as a form of natural science) but, rather, belongs in philosophical, theological, moral, political, jurisprudential, and everyday psychological discourse (along with a good deal of applied psychology discourse as well). In these latter discourses, it is *persons*, not brains, or bodies, or even minds, to whom free agency is ascribed. In sum, new knowledge about the brain, however extensive, leaves the question of free will or free agency where it was before, since the brain is not the kind of thing that can be said to have free agency.

With the legal setting in mind, I think we can conclude that the large issue for debate is not free will versus determinism but the relationship between natural science knowledge (including psychological knowledge that has the form of natural

[22] The literature is very large. I draw here on an earlier work on the history of the insanity defence: Smith, 1981, 1991. I related this to attribution research in Smith, 1985.

[23] There are exceptions. For example, areas of consumer protection law impose strict liability on producers, which means that a defendant is responsible if a certain thing happens, whatever the defendant's agency.

science knowledge) and the kind of discourse that politics and jurisprudence—and everyday psychology and a lot of professional psychology too—articulate. Whether someone has agency in a political or juridical sense is a matter for debate within the languages and social practices of politics and jurisprudence. Of course, there has been and is a place in this for medicine, for psychology, and for social science. The role of these disciplines is to provide clear and accurate descriptions, in terms that doctors, psychologists, and social scientists agree (or try to agree) are the best within their respective fields, of the capacities and illnesses individual people do or do not have. Provided with such descriptions, courts, administrators, and ordinary people are then in the best position to judge whether or not a person (or indeed an institution) has the status that legal rules and custom identify as agency. The decision is not a matter of empirical proof, however much the rhetoric of justification may use empirical language, but of formal or informal rule following by using empirical evidence. It is not the presence of illness in itself that leads to an acquittal in a criminal trial but the convention that a legal system may allow evidence of psychotic illness to signify that a person does not have legal capacity (Morse 2004, 2007).

It is illuminating to draw a parallel with the legal standing of children. The age of a child is part of what makes a child the kind of person who can be said to have agency of the legally appropriate kind, or not. The law has to draw distinctions between ages when a child is and is not responsible in relevant ways. Different jurisdictions have different rules, and this is a matter of social morality and custom not of natural science knowledge. Everyday responses to what children do, operate with highly flexible and negotiable language about the relative agency of children. It is not the age of the child that makes the child free or not free but, rather, the social process of attributing free agency, a social process that takes the growing child to have changing degrees of agency.

A parallel kind of flexibility, allowing for at least some element of negotiation, has become characteristic of the legal administration of mental illness in criminal courts in the past century and a half or so. There are, for example, procedures for pre-trial determination of mental illness and unfitness to plead, there is in England, since 1957 (following Scottish example), the much used plea of diminished responsibility, and there are possibilities for transferring defendants after trial between prison and hospital. All these add up, in principle, to a flexible social means for arbitrating the status of agent and of patient. (How flexible and just it is in practice is not now the point of argument.)

Full of enthusiasm for the advance of physiology, some Victorian medical witnesses in the courtroom made the naïve point that scientific progress, since it had made it possible to describe causal events in the body, enabled experts to say whether a defendant was constrained by illness to do what he or she did and hence to say whether he or she was responsible. Defence lawyers played a substantial part in trying to put such argument before juries (as they have done more recently, for example, in connection with compulsive eating disorders; Eigen 1995). Victorian doctors, however, as they gained experience of giving evidence about insanity and faced hostility when they claimed that their knowledge proved legal incapacity, came to understand the legal context in which they performed. On the occa-

sions when Victorian doctors, citing progress in physiology, claimed that the courts should take much more notice of or even defer to scientific evidence about insanity, judges made two persuasive objections. The first was that if insanity were explicable in causal terms and as a result of *this*, insane actions were not culpable, *all* human conduct would be similarly explicable and not culpable. The idea that no action was culpable was dismissed. Indeed, few people have ever wanted to promote it; the Lombrosian school of criminology is a major exception.[24] The second point was that the issue before the courts was whether the defendant, at the time of the crime, possessed a certain state of mind (under the M'Naghten rules that formalized the matter in1843, this concerned whether the defendant knew what he or she was doing at the time of the crime). While medical evidence about illness did indeed at times help the court decide reach a judgment, the establishment of the presence of illness, even psychotic illness, did not in itself require a finding of non-culpability. As medical witnesses began to understand and accept these points, doctors began to specialize in this kind of work and to acquire their own professional expertise. This was the beginning of the specialty of forensic psychiatry. A new profession, along with new administrative procedures, turned legal decisions about the insanity defence into a regular and ordered process.

A range of contingent factors affected the decisions actually taken. Over time, medical consensus about the existence of certain forms of psychosis usually led courts, where medical opinion agreed in diagnosing psychosis, to transfer the accused to hospital before the court hearing or to accept the insanity defence (or diminished responsibility). Because the courts tried to give understandable evidence for decisions, this could look as if they took the illness itself to be the deciding factor, though, formally speaking, it was not. Controversy lessened with time, especially, in Europe, with the ending of capital punishment, and with the introduction of the verdict of diminished responsibility, a verdict of guilt but guilt to a degree lessened by illness (or, we might say, by constraint). There have been legal parallels to this in laws which make allowance for provocation, self-defence, crimes of passion, or simply actions with unintended consequences (as in manslaughter charges).

Nevertheless, every so often an exceptional case, a case where there is a large and emotional public interest, leads to renewed comment about what is being decided. When I first thought about the history of this area, the cases of Peter Sutcliffe (the English 'Yorkshire ripper' 1981) and of John Hinckley (the US would-be murderer of President Regan, in 1982) did this. The Hinckley case, indeed, led to major changes in Federal and State law. More recently, there is the exceptional case of Anders Behring Breivik. In the media, and it would appear even in the court itself, the decision about reaching a verdict was put as a decision about the finding of

[24] Lombroso and his Italian followers campaigned for—but did not achieve—a root and branch replacement of the legal system by a system for the scientific determination of the type to which a person belongs, along with legislation to provide appropriate therapy/punishment for each type. See Gibson, 2006.

mental illness.[25] The court had the hugely symbolic role, as *the institutional agent* of Norwegian society as a collectivity, of deciding what to do in response to unprecedented violence aimed at the very principles of liberal community. The possibility of finding insanity had an important part in debate about what verdict would be socially and legally acceptable, but the court could not escape its socially assigned agency as the decision-making body. The decision was not only an empirical matter. Perhaps the awkwardness the court exhibited in its role points to why people keep commenting on the insanity defence: in the last analysis, there is no 'right' answer. In decisions that are not routine, there is always an element of ambiguity, the ambiguity of a social world in which people are compatibilists, building social relations on a language which mixes causes and intentions and in which social relations always raise the possibility of questions about responsibility. That is social life in conditions of liberal individualism.

All this lies in the background of the development of forensic psychology. A specialty so labelled developed in the past 50 years or so, though there was, in some jurisdictions (especially in the US), substantial interest in giving psychological evidence a place in the courts early in the twentieth century. The sort of evidence that psychologists had to offer was often about individual capacities or differences in responsiveness to circumstances. For example, psychologists gave evidence about the degree to which witnesses might be considered reliable observers, and psychologists presented evidence about whether there were direct causal links between watching pornographic violence and behaviour. Argument about such evidence did not directly discuss volition and causal determinism but rather sought persuasive, empirically grounded statements about the capacity, or psychological agency, of individuals. The evidence was specifically psychological and enriched description of what a person was thought to be able to do. It gave the courts more to go on in reaching decisions and in ensuring the public found decisions plausible. Some people have thought that neuropsychology should greatly enhance psychology's place in the assessment of agency. But that aspiration has been overtaken by a social and political shift away from concern with assessment of agency (based on knowledge about the past), to assessment of risk (based on predictive knowledge). The courts are now turning to neuroscience evidence in the hope that it will enhance their capacity to estimate risk and hence decide what to do with offenders.

Whatever the utilitarian and administrative arguments for a system of governance based on predictable outcomes and on calculations of risk, rather than governance based on retrospective ascription of agency, contemporary western societies by and large maintain psychological accounts of agency as part of a discourse on responsibility. Jurisprudence has not thrown out the *mens rea* requirement, and public opinion still

[25] There were two determinations, with different medical experts, of the mental health of the accused, and they reached opposite conclusions. The court, in its final judgment that the accused had been a responsible agent, appeared to rely on the empirical evidence that the accused was not mentally ill, or at least not mentally ill in a way and to the degree which exculpates. All the same, the court took a legal decision (also, inescapably, a highly political one) and in so doing used the medical evidence from the second hearing about the accused's mental state to bring the decision within the scope of conventional procedure.

wants punishment for offenders. Everyday psychology continues to be much interested in the psychological states thought to accompany responsibility, and courts are expected to, and do, reflect this. How far this discourse will diminish, remains to be seen.

1.4 Conclusion

There is a substantial body of psychological knowledge that has enriched and rendered precise everyday discursive practices that continuously negotiate the attribution of agency to people. In psychological knowledge with an exclusively behaviourist or neuroscientific form, however, many people would judge the category of agency, as opposed to the category of cause, out of place. Agency is now commonly taken to be a psychosocial category and an attribute of people, or of selves, and, by analogy, institutions, and (logically) not something attributable to brains or bodies. The study of agency is the study of the way people assign, feel, and act on power in all its forms, from desire to governance, or to constraint on power, in their own lives and in the lives of others. The study of agency is hence also the study of how people attribute freedom, obligation, and responsibility. History clearly has a lot to say about this—all the more so as political, legal, and religious thought and practice developed the notion of the person as agent, and, analogously, of institutions as agencies, over many centuries, long before there was a specifically psychological notion of agency.

What I have just stated is muddied, however, by the fact that in earlier centuries 'agency' denoted the powers or capacities of states of existence like souls, gods, and material substances. And this usage, in which one might substitute the word 'cause' (or 'prime cause') for 'agent', continues. Thus, it is not linguistically incorrect to describe neurons, pharmaceuticals, emotions, intelligence, the genes, or whatever as agents. Similarly, one might refer to the agency of the will or of positive thought. But clearly, something different is meant when we refer to a person's agency, and there is a body of opinion that would like to restrict the notion of agency to people—to psychosocial agency. All the same, what I am analytically distinguishing as two conceptions of agency often enough are merged in practice. This is yet another of many examples of the way psychology, taken in all its variety, has a hand in both natural science and in everyday forms of understanding people, and in explanations in terms of material causes and in explanations in terms of intentions, language, and mental processes.

Psychological discussions of agency are embedded in a web of historically formed meanings and power relations. Consider the spiritualist séance. In Victorian and Edwardian times, there was much debate about how to ascribe agency: to spirits, to hysterical women, to charlatans, to conjurors, to women seeking empowerment in a man's world, to nervous reflexes or ideomotor action, to unconscious forces, to extrasensory perception, and so on.[26] It certainly matters to examine what

[26] In general, Oppenheim, 1985; for the agency of women, Owen, 1990; and for the deep questions for psychologists who wish to know what was going on, Lamont, 2013.

went on and what was said about what went on in terms of participants' and observers' understanding of agency. The appropriateness of understanding agency in psychological terms at all (for example, attributing it to unconscious forces) was itself an issue. The most dismissive male doctors were inclined to attribute causation to unstable and even degenerate female bodies. The most ardent devotees were unshakeable in their belief that agency rested with spirits. Worldly sceptics attributed agency to the desire of people to make money. When historians write about the events, they too exercise agency in choosing which story to tell. For historians of psychology, the prime story is the place such phenomena had in the emergence of twentieth-century theories of suggestive influence and of the unconscious. A feminist might think differently.

As the history of the spiritualist séance suggests, the history of notions of psychosocial agency is inseparable from the history of notions of the self. I have used this essay to focus on the notion of volition, rather than on the self, however, and then to use medicolegal discussion to open up a more specific account, relevant to agency, of relations between forms of knowledge and social practice. The history of agency, understood psychosocially, is a history of claims about what individual people could and could not do, and hence of claims about what people could and could not be held responsible for. As early forms of character assessment like physiognomy and phrenology suggest, as fascination with hypnotic and séance states confirms, and as studies of late nineteenth-century and twentieth-century popular psychology bear witness, psychological work was intimately bound up with hopes and fears about what people could or could not do and how individual capacities could or could not change (Hayward 2014; Rose 1989/1999; Thomson 2006). The history of many modern psychological practices, or technologies, is a history of the modern culture of agency—in the work of therapy and counselling, and in the literature and training practices of self-help, memory improvement, acquisition of kills, bolstering of assertiveness, positive thinking, and so on. We may understand academic social psychology as having sought to give these practices a basis in systematic knowledge of interpersonal relations.

Earlier moral philosophers and psychologists, when they assigned agency to people, referred to theoretical entities like states of mind, intentions, volition, and free will, and more recently traits and personality. In the course of the twentieth century, recourse to such entities became a source of division among psychologists. There have been psychologists who decry such references—most recently and emphatically the eliminativist neuropsychologists who held that it was the task of science to translate everyday language about mind (disparagingly called 'folk psychology') into the language of neural events. Needless to say, there were many commentators, psychologists among them, who thought this completely wrong.[27] Moreover, many critics thought a non-eliminativist stance underpinned the ethical principle of respecting people's agency.

[27] The formal statement of the eliminativist argument was given in Churchland, 1981/1989. For reasons why everyday psychology is not to be eliminated, Kusch, 1999, part 3. For a fine statement of opposition to neural reductionism, Gergen, 2010.

All this activity and debate is characteristic of psychological society. As a form of social order, such society has many proponents, both lay and professional, and many detractors. Evaluation often turns on the question of agency. Awareness that agency is a social status, a status ascribed according to rules regulating power, leads critics of psychological society to claim that psychological thought individualizes notions of power and thereby hides its true nature, which in the political worlds we have is a function of socially inequitable structures. By contrast, practicing psychologists—and we may include both those who claim knowledge or skill and those without formal qualifications who seek knowledge or skill—perform with the conviction that individuals can achieve meaningful agency. This exhibits a driving assumption of political individualism: meaningful power and meaningful change exist at the level of the individual. Nevertheless, there are critical psychologists, such as K. J. Gergen, who argue that only due recognition of the social constitution of individual agency will free psychological practice to contribute to rather than diminish individual human agency.[28] Social psychologists, I think, have acknowledged something along these lines, insofar as their research has turned towards the ways in which people understand, ascribe, and negotiate the terms of agency in social relations.

The discourse of specifically personal agency manifestly has high value in contemporary liberal democracies. Linguistic usage going back to earlier centuries, however, makes it possible to describe the body, the brain, the unconscious, the social environment, economic forces, or indeed spirits, God, and devils as agents. Thus, the question, *where* to assign agency, and *to what*, places psychology at the centre of social and political debate about the power of the individual.

References

Arendt, H. (1978). *The life of the mind, Vol. 2, Willing*. London: Secker & Warburg.

Baer, J., Kaufman, J. C., & Baumeister, R. F. (Eds.). (2008). *Are we free? Psychology and free will*. Oxford: Oxford University Press.

Barnes, B. (2000). *Understanding agency: Social theory and responsible action*. London: Sage.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Biran, F., Maine de (2005). *De l'apperception immédiate. Mémoire de Berlin de 1807* (A. Devarieux (Ed.)). Paris: Le Livre de Poche.

Borch-Jacobsen, M., & Shamdasani, S. (2012). *The Freud files: An inquiry into the history of psychoanalysis*. Cambridge: Cambridge University Press.

Carpenter, W. B. (1853). *Principles of human physiology* (4th ed.). London: Churchill.

Carroy, J., Ohayon, A., & Plas R. (2006). *Histoire de la psychologie en France: XIXe-XXe siècles*. Paris: La Découverte.

Chernyshevsky, N. G. (1863/1989). *What is to be done?* (Translated: Katz, M. R., 1989) Ithaca: Cornell University Press.

[28] Gergen's work developed a notion of 'relational being': Gergen, 2009.

Churchland, P. M. (1981/1989). Eliminative materialism and the propositional attitudes. In P. M. Churchland (Ed.), *A neurocomputational perspective: The nature of mind and the structure of science* (pp 1–22). Cambridge: MIT.

Danziger, K. (2001). The unknown Wundt: Drive, apperception, and volition. In R. W. Rieber & D. K. Robinson (Eds.), in collaboration with A. L. Blumenthal and K. Danziger, *Wundt in history: The making of a scientific psychology* (pp. 95–120). New York: Kluwer, Plenum.

Danziger, K. (2008). *Marking the mind: A history of memory*. Cambridge: Cambridge University Press.

Daston, L. (1978). British responses to psycho-physiology, 1860–1900. *Isis,69,* 192–208.

Derkson, M. (2011). The brain as an instrument: Comment on Gergen's 'The acculturated brain'. *Theory & Psychology, 21*(6), 840–845.

Descartes, R. (1649/1985). The passions of the soul. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of Descartes* (Translated: R. Stoothoff, 1985) (Vol. 1, pp. 325–404). Cambridge: Cambridge University Press.

Dostoevsky, F. M. (1865/1960). *Letters from the underworld*. (Translated: C. J. Hogarth, 1960). London: Dent.

Eigen, J. (1995). *Witnessing insanity: Madness and mad-doctors in the English court*. New Haven: Yale University Press.

Erasmus, D., & Luther, M. (1961/2007). *Discourse on free will* (Translated: E. F. Winter, 2007). London: Continuum.

Frank, J. (1986). *Dostoevsky: The stir of liberation, 1860–1865*. Princeton: Princeton University Press.

Frede, M. (2011). *A free will: Origins of the notion in ancient thought*. Berkeley: University of California Press.

Gergen, K. J. (2009). *Relational being: Beyond self and community*. New York: Oxford University Press.

Gergen, K. J. (2010). The acculturated brain. *Theory & Psychology, 20*(6), 795–816.

Gibson, M. S. (2006). Cesare Lombroso and Italian criminology. In P. Becker & R. F. Wetzell (Eds.), *Criminals and their scientists: The history of criminology in international perspective* (pp. 137–158). Cambridge: Cambridge University Press, and Washington D. C.: German Historical Institute.

Goldstein, J. (1994). 'Foucault and the post-revolutionary self: The uses of Cousinian pedagogy in nineteenth-century France'. In J. Goldstein (Ed.), *Foucault and the writing of history* (pp. 99–115). Oxford: Blackwell.

Good, J. M. M. (2000). Disciplining social psychology: A case study of boundary relations in the history of the human sciences. *Journal of the History of the Behavioral Sciences,36,* 383–403.

Hacking, I. (1983). nineteenth-century cracks in the concept of determinism. *Journal of the History of Ideas, 44,* 455–475.

Harris, B. (1979). Whatever happened to Little Albert? *American Psychologist, 34,* 151–160.

Harris, B. (2011). Letting go of Little Albert: Disciplinary memory, history, and the uses of myth. *Journal of the History of the Behavioral Sciences, 47*(1), 1–17.

Hayward, R. (2014). *The transformation of the psyche in British primary care, 1870–1970*. London: Bloomsbury.

Heidegger, M. (1987/2001). *Zollikon Seminars. Protocols—Conversations—Letters*. M. Boss (Ed.), (Translated: F. Mayr & R. Askey). Evanston: Northwestern University Press.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review, 51,* 358–374.

Hewstone, M. (1989). *Causal attribution: From cognitive processes to collective beliefs*. Oxford: Blackwell.

Huxley, T. H. (1874/1894). On the hypothesis that animals are automata and its history. In *Method and results;Essays* (Collected Essays, vol. 1, pp. 119–250). London: Macmillan.

Huxley, T. H. (1893/1989). Evolution and ethics. In J. Paradis & G. C. Williams (Eds.), *Evolution and ethics. T. H. Huxley's evolution and ethics: With new essays on its Victorian and sociobiological context* (facsimile of 1894 edn.). Princeton: Princeton University Press.

Hyslop, J. H. (1894). Freedom, responsibility and punishment. *Mind,* new series *3,* 167–189.
James, W. (1890/1950). *The principles of psychology*. (Reprint, Vol. 2). New York: Dover.
Jaspars, J., Fincham, F. D., & Hewstone, M. (Eds.). (1983). *Attribution theory and research: Conceptual development and social dimensions*. European Monographs in Social Psychology 32. London: European Association of Experimental Social Psychology, Academic Press.
Kane, R. (2005). *A contemporary introduction to free will*. Oxford: Oxford University Press.
Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.
Kelley, H. H. (1971). Moral evaluation. *American Psychologist, 26,* 293–300.
Kimble, G. A., & Perlmuter, Lawrence C. (1970). The problem of volition. *Psychological Review, 77,* 361–384.
Kruks, S. (2012). *Simone de Beauvoir and the politics of ambiguity*. New York: Oxford University Press.
Kusch, M. (1999). *Psychological knowledge: A social history and philosophy*. London: Routledge.
Lamont, P. (2013). *Extraordinary beliefs: A historical approach to a psychological problem*. Cambridge: Cambridge University Press.
Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
Law, J., & Hassard, J. (Eds.). (1999). *Actor network theory and after*. Oxford: Sociological Review Monographs, Blackwell.
Lloyd, G. E. R. (2007). *Cognitive variations: Reflections on the unity and diversity of the human mind*. Oxford: Oxford University Press.
Lloyd, G. E. R. (2009). *Disciplines in the making: Cross-cultural perspectives on elites, learning, and innovation*. Oxford: Oxford University Press.
McClure, J. (2012). Attributions, causes, and actions: Is the consciousness of will a perceptual illusion? *Theory & Psychology, 22*(4), 402–419.
Michotte, A., & Prüm, E. (1911). Étude expérimentale sur le choix voluntaire et ses antécédents immédiats. *Archives de Psychologie, 10*, 113–320.
Mill, J. S. (1843/1900). *A system of logic ratiocinative and inductive*. (8th ed., reprint). London: Longmans, Green.
Morse, S. J. (2004). New neuroscience, old problems. In B. Garland (Ed.), *Neuroscience and the law: Brain, mind, and the scales of justice* (pp. 157–198). New York : Dana, American Association for the Advancement of Science.
Morse, S. J. (2007). The non-problem of free will in forensic psychiatry and psychology. *Behavioral Sciences and the Law, 25,* 203–220.
Oppenheim, J. (1985). *The other world: Spiritualism and psychical research in England, 1850–1914*. Cambridge: Cambridge University Press.
Owen, A. (1990). *The darkened room: Women, power, and spiritualism in late Victorian England*. Philadelphia: University of Pennsylvania Press.
Pinkard, T. (2002). *German philosophy 1760–1860: The legacy of idealism*. Cambridge: Cambridge University Press.
Pockett, S., Banks, W. P., & Gallagher, S. (Eds.). (2006). *Does consciousness cause behavior?* Cambridge: MIT.
Rodder, H., & Meynen, G. (2013). Does the brain "initiate" freely willed processes? A philosophy of science critique of Libet-type experiments and their interpretation. *Theory & Psychology, 23*(1), 3–21.
Rose, N. (1985). *The psychological complex: Social regulation and the psychology of the individual*. London: Routledge & Kegan Paul.
Rose, N. (1989/1999). *Governing the soul: The shaping of the private self*. (2nd ed.). London: Free Association.
Rose, N. (1996/1998). *Inventing our selves: Psychology, power, and personhood*. Cambridge: Cambridge University Press.
Rose, N. (2007). *The politics of life itself: Biomedicine, power, and subjectivity in the twenty-first century*. Princeton: Princeton University Press.

Rose, N. (2013). The human sciences in a biological age. *Theory, Culture & Society, 30*(1), 3–34.
Ryle, G. (1949/1963). *The concept of mind*. Harmondsworth: Penguin.
Sechenov, I. M. (1935/1968). Reflexes of the brain. In *Selected works*. (Translated: A. A. Subkov, facsimile reprint, pp. 263–336). Amsterdam: E. J. Bonset.
Sechenov, I. M. (1965). *Autobiographical notes*. Donald B. Lindsley (Ed.). Washington: American Institute of Biological Sciences. (Translated: Kristan Hanes, written about 1904).
Seigel, J. (2005). *The idea of the self: Thought and experience in Western Europe since the seventeenth century*. Cambridge: Cambridge University Press.
Shamdasani, S. (2003). *Jung and the making of modern psychology: The dream of a science*. Cambridge: Cambridge University Press.
Shuttleworth, S. (2010). *The mind of the child: Child development in literature, science, and medicine, 1840–1900*. Oxford: Oxford University Press.
Sirotkina, I. (2002). A family discussion: the Herzens and the science of man. *History of the Human Sciences, 15*(4), 1–18.
Sirotkina, I., & Smith, R. (2012). Russian Federation. In D. B. Baker (Ed.), *The Oxford handbook of the history of psychology: Global perspectives* (pp. 412–441). Oxford: Oxford University Press.
Smith, R. (1981). *Trial by medicine: Insanity and responsibility in Victorian trials*. Edinburgh: Edinburgh University Press.
Smith, R. (1985). Expertise and causal attribution in deciding between crime and mental disorder. *Social Studies of Science*, *15*, 67–98.
Smith, R. (1991). The law and insanity in Great Britain, with comments on Continental Europe. In F. Koenraadt (Ed.), *Ziek of Schuldig? Twee eeuwen forensiche Psychiatrie en Psychologie*. (pp. 247–281). Arnhem: Gouda Quint, and Amsterdam: Rodopi, for Willem Pompe Instituut voor Strafrechtswetenschappen te Utrecht.
Smith, R. (1997). *The Fontana history of the human sciences*. London: Fontana. Also published as *The Norton History of the Human Sciences*. New York: Norton.
Smith, R. (2005). The history of psychological categories. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36,* 55–94.
Smith, R. (2011). 'The sixth sense': Towards a history of muscular sensation. *Gesnerus: Swiss Journal of the History of Medicine and Sciences, 68,* 218–271.
Smith, R. (2013a). *Free will and the human sciences in Britain, 1870–1910*. London: Pickering & Chatto.
Smith, R. (2013b). *Between mind and nature: A history of psychology*. London: Reaktion.
Tallis, R. (2011). *Aping mankind: Neuromania, darwinitis and the misrepresentation of humanity*. Durham: Acumen.
Tauber, A. I. (2010). *Freud, the reluctant philosopher*. Princeton: Princeton University Press.
Tauber, A. I. (2013). Freud's social theory: Modernist and postmodernist revisions. *History of the Human Sciences, 25*(4), 43–72.
Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Cambridge: Cambridge University Press.
Thomson, M. (2006). *Psychological subjects: Identity, culture, and health in twentieth-century Britain*. Oxford: Oxford University Press.
Turgenev, I. S. (1862/1960) Fathers and sons. *The Vintage Turgenev*. New York: Vintage Books (Vol. 1, pp. 165–352, translated: Harry Stevens, 1960). [The original Russian title translates as *Fathers and Children*].
Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge: MIT.
Wyhe, J. van (2004). *Phrenology and the origins of Victorian scientific naturalism*. Aldershot: Ashgate.

Chapter 2
The Providence of Associated Minds: Agency in the Thought of Giambattista Vico and the Origins of Social and Cultural Psychology

Luca Tateo

I would like to tackle the concept of agency in psychology from the particular perspective of the seventeenth-to-eighteenth-century scholar Giambattista Vico. Vico (1668–1744) was indeed a philosopher, rhetorician, historian, and jurist from the Kingdom of Naples, South Italy (Fig. 2.1), whose influence has been fundamental, though sometimes neglected, for the development of some ideas that became part of the legacy of social and cultural psychology. His *opus magnum* was *The New Science* (Vico 1948), subtitled in its third and final edition originally published in 1744: "about the common nature of nations". This work was the final leg of an intellectual journey, during which Vico tried to build an innovative project of "a rational civil theology of divine providence" (Vico 1948, p. 4). At that time, it was commonly understood that studying nature in order to admire and praise the work of God through its creation was a legitimate task of science. This natural theology argument enabled the full deployment of natural sciences, but also created a divide between studying nature and human beings. Natural laws were indeed subject only to the will of God, and these manifestations were readable because they were written in the language of mathematics, like Galilei stated (Galilei 1960). But the laws governing human actions were a different matter. Between human beings and God, there were two relevant open issues: the original sin and the free will.

As Smith (Chap. 1, this volume) argues, accounting for the reasons of human behavior was a realm of metaphysics. Human beings are not mechanically obeying to the laws of nature, they are rather likely to violate them, for the good or for the bad, and the need for accounting for the violation was the reason for the birth of psychology, as part of the metaphysics, precisely.

Giambattista Vico aimed at developing an original and all-embracing solution to the problem of explaining the relationship between human mind, civilization, and divine design. His answer was to develop a whole philosophical system that has been one of the grounds for the development of social and cultural psychology, as I will argue in the following sections. The reason for Vico's new way of understand-

L. Tateo (✉)
Niels Bohr Centre for Cultural Psychology, Aalborg University, Aalborg, Denmark
e-mail: luca@hum.aau.dk

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_2

Fig. 2.1 The statue of Vico in the center of Naples

ing the relationship between mind and culture being extremely innovative at that time is that he conceptualized the complex process of co-development of mankind as historically situated, language based, and activity based. The originality of Vico was that of making the products of human activity based on language—such as art, law, myths, religion, etc.—an acceptable object of science (Burke 1985). All human products of social and practical life in general that could be referred to the domain of language were expelled from scientific discourse by Cartesianism (Berlin 1974), to the extent that they could not be understood and represented in mathematical terms. Cartesianism indeed rejected all the forms of knowledge related to the use of language, such as rhetoric, as mere forms of presentation and orientation of beliefs. Vico instead "hoped to find almost a compendium of that inductive method which he attempted to 'transfer from natural things to human and civil things'" (de Mas and Houck 1971, p. 90). The general principles of his view are

> (a) that human nature and society are not fixed or stationary, but rather are in a state of continuous change; (b) that the changes occur in evolutionary cycles influenced by human events; (c) that despite epistemological limitations, it is scientifically possible to investigate social behavior across eras in order to reveal events that influence the recursive evolution of society, as well as the genesis of theories of human behavior and society. (Rosnow 1978, p. 1322)

2.1 The New Science

The intellectual project that Vico pursued all his life was that of accounting for the whole history of human development of civilization as well as to the full range of human products, arts, law, customs, language, institutions, etc. This scientific

enterprise was possible because the object of the "new science" is made by the humans themselves. "Now, as geometry, when it constructs the world of quantity out of its elements, or contemplates that world, is creating it for itself, just so does our Science, but with a reality greater by just so much as the institutions having to do with human affairs are more real than points, lines, surfaces, and figures are" (Vico 1948, pp. 104–105). From the metaphysical point of view, instead, Vico aimed at reconciling the immanent role of divine providence with the historical development of civilization and the richness of human psychology. Therefore, he developed an anthropological metaphysics, whose units of analysis were the complex products of human *ingegno* rather than the atomistic and self-referential concept of Descartes. "Vico's active epistemology presupposes an internal relationship between creators (that is, God and man), creation, and knowledge. He who has created something can know what he has created. God created the 'world of nature' and therefore He alone can truly know this world; man, on the other hand, created and knows the two 'worlds' of 'quantity' and of 'nations'" (Tristam 1983, p. 148). Thus, a central role is played by human agency, whose distinctive characteristics are will and intellect. Will is "the property of human nature which not even God can take from man without destroying him" (Vico 1948, p. 109)—that is the foundation for his active production of his own world:

> This authority is the free use of the will, the intellect on the other hand being a passive power subject to truth. For from this first point of all human things, men began to exercise the freedom of the human will to hold in check the motions of the body, either to subdue them entirely or to give them better direction (this being the impulse proper to free agents, as we have said above in the Method). (Vico 1948, p. 109)

Vico introduces a fundamental difference in philosophical anthropology. In fact, he conceptualizes the history of civilization as progressive development of the relationship between will and intellect, with the latter taking over the former. Almost in a Vygotskian way, Vico argues that the distinctive feature of human nature has been the capability of creating products of civilization—namely divinity worship, marriage, and burials—as self-regulatory systems that were able to act "on the bestial passions" of primitive men and "transformed them into human passions" (Vico 1948, p. 90). The study of this historical and psychological collective process forms a specific object, for which "this Science must therefore be a rational civil theology of divine providence" (Vico 1948, p. 90). The role of divine providence is in fact that of guiding, setting the conditions for the survival of the human race. Beyond that, civilization follows its own multiple pathways, generating the "world of nations in all the extent of its places, times and varieties" (Vico 1948, p. 92). Through this innovative conceptual operation, Vico carved out a specific space for a science of human activity that became autonomous with respect to both theology and natural philosophy:

> Our Science is therefore a history of human ideas, on which it seems the metaphysics of the human mind must proceed. This queen of the sciences, by the axiom that the sciences must begin where their subject matters began took its start when the first men began to think humanly, and not when the philosophers began to reflect on human ideas. (Vico 1948, p. 92)

2.2 Homo Faber

The metaphysical view about human agency that Vico outlines constitutes the framework for his anthropological and psychological concept of agency. "Vico finds the underlying basis of this pattern in a metaphysics of the human mind" (Pompa 2002, p. xxvi). The principles of this view of agency are (a) that human beings are essentially imaginative and poietic creatures, (b) the interdependence of empirical and rational, (c) the historical and collective dimension of human behavior, and (d) the central role of language in creating human reality. I argue that these principles will be elaborated by several scholars in the eighteenth century, leading to the development of social and cultural psychology.

2.2.1 Imaginative Function

The first point concerns Vico's theory of human psyche. In fact, he found the distinctive characteristic of human psyche in its capability of imagination. With a hazardous anachronism, I would say that imagination in Vico is what we call today symbolic capability. In his own words, indeed, imagination "is nothing but the springing up again of reminiscences, and ingenuity or invention is nothing but the working over of what is remembered" (Vico 1948, p. 236). It is a progressive distancing from the senses through the creation of images that allows the construction of abstract concepts. Imaginative capability is based on three fundamental functions of the mind: *fantasia*, the capability to imitate and change; *ingegno,* the capability to create correspondence between things; and *memoria,* the capability to remember. Nevertheless, there is a fundamental anthropological difference between primitive men, who own these capabilities as a result of the divine providence farsightedness, and contemporary human beings. In fact, in the course of historical civilization, humans also develop the function of rational thinking:

> Now, since the human mind at the time we are considering had not been refined by any art of writing nor spiritualized by any practice of reckoning or reasoning, and had not developed its powers of abstraction by the many abstract terms in which languages now abound, as we said above in the Method, it exercised all its force in these three excellent faculties which came to it from the body. All three appertain to the primary operation of the mind whose regulating art is topics, just as the regulating art of the second operation of the mind is criticism; and as the latter is the art of judging, so the former is the art of inventing, as has been said above in the last Corollaries of the Poetic Logic. And since naturally the discovery or invention of things comes before criticism of them, it was fitting that the infancy of the world should concern itself with the first operation of the human mind, for the world then had need of all inventions for the necessities and utilities of life, all of which had been provided before the philosophers appeared (Vico 1948, p. 236).

In this sense, primitive and modern minds are incommensurable, thought the imaginative capability has not disappeared but became a legacy of mankind that has been educated, but also weakened, by the development of rational thinking in the history of civilization. Thus, we cannot have full access to the mind of primitive peoples, but indirectly through the study of the products of their mind, mainly language, myths, and art.

2.2.2 Empirical and Rational

Following Francis Bacon, Giambattista Vico always considers the theoretical and practical dimensions of knowledge in relationship. Knowledge is made of understanding things as they came to be as they are at present. As human cognition is limited in his capability to know the real world, only God has the full understanding of phenomena and their causes, as He made them and knows the whole history. "But this contrast also implies a parallel. God knows (*cognoscit*) because he creates and disposes; man knows (*novit*) because he makes and composes. The active component of human knowing is the key to man's participation in the divine form of cognition, *intelligere*" (Barnouw 1980, p. 616). Since the origins of civilization, indeed, knowledge has been related to action, or making, to the extent that the primary requirement of any form of knowledge was the survival of the individuals and their kin in the wild nature. Vico calls this dawning of civilization the poetic age, understanding poetic in the twofold sense of its Greek etymology "Poïesis" (Ancient Greek: ποίησις), deriving from the verb ποιέω, which means "to make," and of the imaginative work of poetry, which was the first form of structured linguistic knowledge of ancient civilizations.

Vico's theory of knowledge was first formulated in his 1710 work *The Most Ancient Wisdom of the Italians* (Vico 1988), and fully developed in *The New Science.* In the former, he stated that, in ancient Latin, the words *verum* (the true) and *factum* (the created) were interchangeable. The two words had similar meanings which Vico understands as "to know" and "to make." Thus, the knowledge has its object in the products of human activity, at least the form of knowledge that can be attained by human beings. As "science consists in a knowledge of the genesis of things" (Vico 1988, p. 248), the only object whose origins are knowable is the product of human activity itself. There are indeed different kinds of truth: (a) the "truth" (*verum*), which only pertains to God; (b) the "common sense" (*verum certum*) which is the practical knowledge and belief achieved through practices and consent; and (c) the "truth through making" (*verum factum*), which is the scientific knowledge about all the products of human activity.

We are now know all the elements of the complex theory of knowledge elaborated by Vico (Fig. 2.2).

Knowledge originates from the faculty of sense, which at psychological level corresponds to the elaboration of perception. This first material, which is elaborated at the pragmatic level of language, constitutes the experiential fuel for the mind's faculties of *fantasia*, *ingegno,* and *memoria*:

> The human mind is naturally inclined by the senses to see itself externally in the body, and only with great difficulty does it come to attend to itself by means of reflection. This axiom gives us the universal principle of etymology in all languages: words are carried over from bodies and from the properties of bodies to express the things of the mind and spirit. (Vico 1948, p. 70)

The mental activity of connecting, recollecting, and elaborating experiences is progressively crystallized in language. Primarily, it takes the form of metaphor, which, according to Vico, is an elliptical and condensate mythical image. Metaphor is also

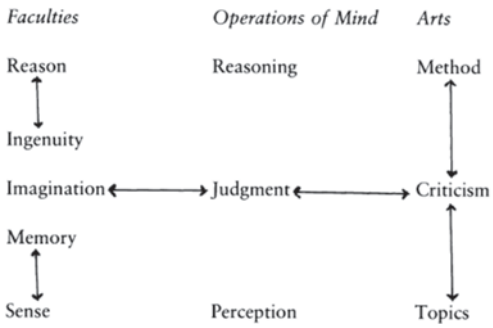

Fig. 2.2 Elements in the production and assessment of knowledge in Vico. (Tristam 1988, p. 360)

the primary form in which knowledge circulates among collectivities. It does not require analytical skills, as its linguistic and iconic form allows an immediate and total apprehension of its meaning. For instance, when we say "motherland," there is no need for exploding the full sense of the word because the affective, historical, and experiential meaning of the metaphor is immediately graspable by anyone, producing a quite predictable effect in any listener, no matter which nation he/she belongs to:

> Men at first feel without observing, then they observe with a troubled and agitated spirit, finally they reflect with a clear mind. This axiom is the principle of the poetic sentences, which are formed with senses of passions and affections, in contrast with philosophic sentences, which are formed by reflection and reasoning. The more the latter rise toward universals, the closer they approach the truth; the more the former take hold of particulars, the more certain they become. (Vico 1948, pp. 67–68)

But this elaboration is made possible by the framing of the culture that provides not only a repertoire of metaphoric images but also the guide for the anticipation of judgment about new experiences: this is the fundamental idea of common sense:

> Human choice, by its nature most uncertain, is made certain and determined by the common sense of men with respect to human needs or utilities, which are the two origins of the natural law of nations. (…). Common sense is judgment without reflection, shared by an entire class, an entire people, an entire nation, or the whole human race. (Vico 1948, p. 57)

It is only through this complex process of elaboration and progressive abstraction that human beings can attain a level of rational reflection. Nevertheless, Vico is very clear in stating that logical reasoning, even though is the highest form of mental activity, is firmly grounded on the other faculties:

> That is, the human mind does not understand anything of which it has had no previous impression (which our modern metaphysicians call 'occasion') from the senses. Now the mind uses the intellect when, from something it senses, it gathers something which does not fall under the senses; and this is the proper meaning of the Latin verb intelligere. (Vico 1948, p. 98)

2.2.3 Historical and Collective Dimensions

The theory of agency elaborated by Vico is based on the idea that primitive human nature was moved by a survival instinct for the preservation of the kin. Progressively, during the conglomeration of primitive families, "must have sprung the impulse (*conato*) proper to the human will, to hold in check the motions impressed on the mind by the body, so as either to quiet them altogether, as becomes the sage, or at least to direct them to better use, as becomes the civil man" (Vico 1948, p. 90). Thus, Vico identifies a third type of agency beside divine providence and individual agency. There is a form of collective agency, whose expression is common sense, that is oriented toward the construction of social organizations that crystallize around the three primordial poetic institutions: worship, marriage, and burials. These forms of collective organization are directly based on the mind's psychological functions, that is why they are common to all the historical forms of civilization. But at the same time, they produced a wide variety of historically and geographically situated variations, according to the concrete conditions in which they developed. "Whenever the time and fashion is thus and so, such and not otherwise are the things that come into being" (Vico 1948, p. 58).

The impulse (*conato*) to form collective organizations is the link between the will of individual agents and the formation of collective bodies. According to Vico, these organizations followed general laws of historical development and modes of thought. The first stage, called the age of gods, was characterized by the poetic logic, in which an undifferentiated fear of natural phenomena led to the creation of anthropomorphic divinities as explanation. The first universal poetic character was Jupiter, as a personification of the thunder. Once this image is created as a form of explanation, it becomes a shared category on which primitive collective nomad groups self-regulate their mutual behavior. The second stage, the age of heroes, is characterized by the stabilization of collective groups in a given territory, in which some prominent families take control in the form of oligarchies. They aggregate larger groups of people to whom they offer protection and safety as rulers in change of subjection. The cultural forms that characterize this period are those of semi-divine heroes that emerge as intermediate figures between the divinity and the laymen. At this stage, the typical poetic universal character is that of Hercules. Finally, in a third stage, the people that were formerly subjugated take the initiative of demanding for the political power and the equal rights over the oligarchy. This is the stage in which the poetic forms of culture are overcome by a more prosaic and vulgar form of language, which is typical of the democracies. In this process, the collective agency is progressively moving from an external and hyperuranic agent to collectives of people. At the same time, the role of imaginative and poetic thinking is decreasing while the reasoning mode of thought is emerging.

This account of the development of collective agency, directly related to the development of different modes of thought and different linguistic forms, will constitute one of the theoretical grounds for the social and cultural psychology that will be born a century later.

2.2.4 Language

Another fundamental topic of Vico's new science is the relationship between language, culture, and mind. Language, thought, and civilization are strictly related. Even though Vico does not overlap language with alphabetic languages, he understands language as a wider symbolic capability that originates from mute and bodily communication. Thus, language includes images, hieroglyphics—that is iconic writing—and finally, alphabetic writing. Vico states that psychosocial processes are crystallized in language, which is the vehicle of cultural continuity. It is also the primary object of investigation for an historical and developmental science of civilization:

> Vulgar traditions must have had public grounds of truth, by virtue of which they came into being and were preserved by entire peoples over long periods of time. It will be another great labor of this Science to recover these grounds of truth which, in the passage of years and the changes in languages and customs, has come down to us enveloped in falsehood. (Vico 1948, p. 58)

Thus, language study, or philology, is the only scientific tool we have to investigate the mind of ancients. The study of metaphors is, for instance, an example of how we can reconstruct the psychological processes that led to the creation of common sense concepts (Danesi 1995).

Language has also fundamental cognitive, ethical, and social functions in the development of civilizations. As a professor of eloquence at the University of Naples, Vico was perfectly aware of the pragmatic and creative role of language. Following again Bacon, Vico develops the idea that language is important not only as an analytical tool but also as a heuristic tool for "the invention of arguments designed to investigate the matter at hand" (Perkinson 1962, p. 35). The cognitive function of language is then to operate as a tool of *critica*—that is, after Descartes, the function of rigorous analysis of knowledge in order to seek the truth but also as a tool for *topica*—that is the development of arguments that "function to uncover new knowledge pertinent to the question in hand" (Perkinson 1962, p. 35). For Vico, language has then an inherent abductive power, as in the case of metaphor, which is fundamental for the advancement of knowledge. "Today, *critica* exclusively is cultivated: *topica*, far from being placed first in order, is completely forced out. And this is wrong, since, as the invention of arguments precedes by nature the evaluation of truth, so *topica* should precede *critica*" (Vico 1965, p. 178).

The ethical and social function is related to the agency of the individual operating in the collectivity. In his work, "De nostri temporibus studiorum ratione" (On the methods of study of our times; Vico 1965), originally published in 1708, Vico outlines his theory of the specificity of human sciences, polemizing with Cartesianism-based mainstream approach. Vico claims that human thought is based on several dimensions, not just on logical and rational thought, the *critica*. This implies that an education, which is only aimed at developing this function, will grow students that are not able to play an active and constructive role and lead in civil society. "Those who have been taught only *critica* are unable to share with or to teach to the

rest of the community whatever new truths they might obtain. This is because those who are not exercised in *topica* 'never have the experience of immediately seeing whatever persuasive is implicit in every cause'" (Perkinson 1962, p. 37).

Vico's conception of language as a fundamental object of study, to the extent that it embodies the system of knowledge, beliefs, ethics, and history of each civilization, was a fundamental turn in the history of human sciences. Indeed, it opens the possibility of considering the products of art, mythology, and folklore as relevant topics for understanding the modes of thought of a specific culture. This had a direct influence on the development of folklore studies in Romanticism and later on the development of anthropology and folk psychology (Berlin 1976; Danesi 1995; Leach 1976; Diamond 1977). The principles of Vico's human science, using again with an educated anachronism, can be summarized in the idea that mind and civilization are co-constitutive. Human beings are active agents creating their own world. The modes of thought are framed within the forms of civilization that they have contributed to create through collective action under specific historical conditions. The modes of thought and the forms of civilization are connected by language and products of art, who register the development of both. Finally, both modes of thought and forms of civilization follow general laws of historical development that can be studied with a specific method.

2.3 Cattaneo, Wundt, and the Origins of Social Psychology

It is not worth here trying to reconstruct historically and philologically the direct influence of Vico's ideas on psychological sciences. This should be the object of a specific work that I eagerly look forward to see. Such influence is often under track or even neglected and it deserves an appreciation. I would instead try to discuss how the visionary ideas of Vico at that time anticipated the thought of social and cultural psychology a century later. In particular, I will focus on two scholars: Carlo Cattaneo and Wilhelm Wundt. The former is important for being the one creating the term "social psychology" as we today understand it (Tateo and Iannaccone 2011), the latter for being conventionally acknowledged as the founder of scientific psychology.

Carlo Cattaneo was a nineteenth-century Italian philosopher and politician who explicitly draw inspiration from Vico in discussing the development of culture in relation to psychological processes. He focused on the dynamics of continuity and change in society, in that he identified in both endogenous social interactions within a culture and the exogenous interaction between different cultures that assure the creation of new ideas through the contribution of *associated minds* (Cattaneo 2000). Discussing the development of civilization, similar to the one that will be later presented by Wundt (1952), Cattaneo argued that primitive men could only develop individual and limited experience of his world. The spring of civilization and culture activated a process of social construction of knowledge that led to more articu-

lated understandings of the reality, even those that were not directly accessible to the individual experience. Like Vico, Cattaneo understood the mutual evolution of individual mind and culture as a progressive expansion of the sphere of knowledge through the artifacts that were collectively constructed, accumulated, and transmitted in everyday human activity. The construction of new knowledge occurs by the process of "*antithesis*" (Cattaneo 2000, p. 77). Collective life is indeed the context into which individuals confront their points of view and their opposite ideas. Such confrontation generates a positive conflict allowing the improvement of knowledge and by the development of the cultural and material tools—language, technology, means of transport, weapons, memory supports, etc.—allowing to widen the horizons of experience triggering the development of new modes of thought and new activities. Cattaneo claimed that the study of the relationship between mind and culture should be the object of a specific science, that he first called *psicologia sociale* (social psychology; Cattaneo 1964).

According to Cattaneo, the developmental process of culture can be generated by two different mechanisms. The first is the appearance of the "genius," in Vico's sense: individuals capable to turn the experience of the world into discovery. The new ideas are elaborated within the society and become collective. The second mechanism is the collective praxis of "the common people, unaware of academic debates but confident in their capacities and aspirations for better life prospects (…) posing anew, and agitating to resolve, fundamental issues in organized existence" (Sabetti 2006, p. 10). Cattaneo defines these everyday actors of the cultural development as "obscure Socrates" (1960, p. 281). Cattaneo draws on Vico's anthropological philosophy of human agency based on the relationship between "being, becoming and acting" (Sabetti 2006, p. 15). The relationship between individuals and society is again co-constitutive: "society not only *sees* thing, but also *makes* things" (Cattaneo 2000, p. 84, original bold).

Cattaneo also stresses the relationship between individual and collective memory, that already Vico discussed as a fundamental tie between individual development and history of civilization. The work of associated minds allows the creation of trans-generational ties and feelings of common belonging which constitutes the cultural unity. "Society is in possession of all the aids of the artificial memory" (Cattaneo 2000, p. 111)—texts, monuments, images, national symbols, etc.—enabling to overcome the limits of the individual memory and to create a continuity between generations and a cumulative knowledge through the "collective memory, which is the contribution of all the individual memories" (Cattaneo 2000, p. 113). A similar idea is developed by Wundt who considers the collective representations "mental products which are created by a community of human life and are, therefore, inexplicable in terms merely of individual consciousness, since they presuppose the reciprocal action of many" (Wundt 1916, p. 3). In the middle of the nineteenth century, a new way of understanding the relationship between mind and culture: "Vico and later Wundt's demands to turn to culture (myths, language and traditions) and history if fact put modern, nomothetic psychology into a quandary, which Boesch (1971) formulated nicely when he said 'It is the dilemma of psychology that it deals with an object that creates history' (p. 9)" (Eckenseberger 2011, p. 416).

Wundt's account of the development of civilization is very close to Vico and Cattaneo. The three scholars totally agree on the fundamental role of the symbolic forms in the relationship between mind and culture. For Wundt, the products of culture assume above all the form of the alphabetic writing. Starting from the requirements of trade and lawmaking, the system of writing was developed with the purpose of sharing laws and recording economic exchanges. "In this wise, the material aspects of the world culture exerted an influence upon the mental aspects, whose direct expressions are speech and writing" (Wundt 1916, p. 486). The dialectic between mind and culture is a historical process based on collective activities, related to the specific material conditions of a civilization in a particular moment of its development. Also, Wundt realized that the study of the relationship between individual and collective action and between mind and civilization would require specific theories and methods, which he called *Völkerpsychologie*. It is not the case that Wundt's plan of his ten-volume work contains all the topics that Vico covered in his *New Science: Language, Art, Myth and Religion, Society, Right, Culture in History*.

2.4 Conclusions

In this chapter, I have presented an overview of the original contribution of Giambattista Vico to the study of the relationship between mind and culture. His theory also includes an innovative concept of the relationship between individual and collective agency. His innovative ideas originated a new way of looking at mind–culture relationships, at the role of symbolic forms and at the relationship between experience and language, that inspired not only social and cultural psychology but also pragmatism (Fisch 1969), cognitive sciences (Danesi 1995), and hermeneutics (Schaeffer 1987). Vico's intuition was that human phenomena are specific and historically situated, requiring a new theory and method to be understood. This idea is based on his philosophical anthropology and his theory of individual and collective agency, which he tried to organize in a specific branch of knowledge that he proudly called "New Science." After a century, Cattaneo's "social psychology" and Wundt's "Völkerpsychologie" seem to share his idea that human beings are essentially social and "symbolic" creatures, inclined to create his own social environment. Vico's intuition was to understand the relationship between mind and culture in terms of the interweaving between individual and collective agency. Forerunning the development of modern social sciences, Vico looked at the "pervasive pattern which characterizes all the activities of any given society: a common style reflected in the thought, the arts, the social institutions, the language, the way of life and action, of an entire society" (Berlin 1976, p. xvii). Such environment created along the phylogenetic development determines in return the features of the individual psychological processes. In Cattaneo's own words: "*The most social act of men is thinking*" (2000, p. 89). Both Cattaneo and Wundt draw on the idea that this historical process must be studied with specific approaches that observe the most elaborated products of human activities, such as technology, arts, language. Vico introduced the idea that

> this succession is intelligible, and not merely causal, since the relationship of one phase of a culture or historical development to another is not that of mechanical cause and effect, but, being due to the purposive activity of men, designed to satisfy needs, desires, ambitions (the very realization of which generates new needs, desires, ambitions), is intelligible to those who possess a sufficient degree of self-awareness, and occurs in an order which is neither fortuitous nor mechanically determined, but flows from elements and forms of life, explicable solely in terms of human goal-directed activity (Berlin 1976, pp. xvii-xviii).

Acknowledgments This work has been funded by the Marie Curie IEF−2012 grant "EPICS. Epistemology in psychological science, the heritage of Giambattista Vico and the cultural psychology" at Aalborg University (Denmark).

References

Barnouw, J. (1980). Vico and the Continuity of Science: The Relation of His Epistemology to Bacon and Hobbes. *Isis*, 71(4), 609–620.

Berlin, I. (1974). The divorce between the sciences and the humanities. *Salmagundi, 27,* 9–39.

Berlin, I. (1976). *Vico and herder*. *Two studies in the history of ideas*. New York: Chatto and Windus.

Boesch, E. E. (1971). *Zwischen zwei Wirklichkeiten. Prolegomena zu einer ökologischen Psychologie*. Bern: Huber.

Burke, P. (1985). *Vico*. Oxford: Oxford University Press.

Cattaneo, C. (1960). Filosofia della rivoluzione. In N. Bobbio (Ed.), *Scritti filosofici, I* (pp. 272–286). Florence: Le Monnier.

Cattaneo, C. (1864). Dell'antitesi come metodo di psicologia sociale. *Il politecnico, 20,* 262–270.

Cattaneo, C. (2000). *Psicologia delle menti associate*. Roma: Editori Riuniti.

Danesi, M. (1995). *Giambattista vico and the cognitive science enterprise*. New York: Peter Lang.

de Mas, E., & Houck, J. K. (1971). On the new method of a new science: A study of Giambattista Vico. *Journal of the History of Ideas, 32*(1), 85–94.

Diamond, S. (1977). On reading Vico. *Dialectical Anthropology, 2*(1), 19–32.

Eckensberger, L. H. (2011). Cross-cultural differences as meaning systems. In F. J. R. van de Vijver, A. Chasiotis, & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology*. Tilburg: Cambridge University Press.

Fisch, M. (1969). Vico and pragmatism. In G. Tagliacozzo (Ed.), *Giambattista Vico. An international symposium* (pp. 416–417). Baltimore: The Johns Hopkins University Press.

Galilei, G. (1960). The Assayer (English trans: S. Drake and C. D. O'Malley). In S. Drake & C. D. O'Malley (Eds.), *The controversy on the comets of 1618*. Philadelphia: University of Pennsylvania Press.

Leach, E. (1976). Vico and the future of anthropology. *Social Research, 43*(4), 807–817.

Perkinson, H. J. (1962). Giambattista Vico and "The Method of Studies in Our Times": A criticism of descartes' influence on modern education. *History of Education Quarterly,* 2(1), pp. 30–46.

Pompa, L. (2002). Introduction. In L. Pompa (Ed.), G. Vico. The first new science (pp. xix–xxx). Birmingham: Cambridge University Press.

Rosnow, R. L. (1978). The prophetic vision of Giambattista Vico: Implications for the state of social psychology theory. *Journal of Personality and Social Psychology, 36*(11), 1322–1331.

Sabetti, F. (2006). *A quiet revolution: Rethinking the foundations of human society*. Paper for conference on Vincent Ostrom: The quest to understand human affairs, workshop in political theory and policy analysis, Indiana University, Bloomington, Indiana, May 31-June 3, 2006.

Schaeffer, J. D. (1987). Sensus communis in Vico and Gadamer. *New Vico Studies, 5,* 117–130.

Tateo, L., & Iannaccone, A. (2011). Social representations, individual and collective mind: A study of Wundt, Cattaneo and Moscovici. *Integrative Psychological and Behavioral Science, 46*(1), 57–69.

Tristam, R. J. (1983). Explanation in the new science: On Vico's contribution to scientific socio-historical thought. *History and Theory, 22*(2), 146–177.

Tristam, R. J. (1988). Vico on the production and assessment of knowledge. *Philosophy and Phenomenological Research, 48*(3), 355–388.

Vico, G. (1948). *The new science of Giambattista Vico* (trans: T. Goddard Bergin & M. H. Fisch). Ithaca: Cornell UP.

Vico, G. (1965). *On the study methods of our time* (trans: E. Gianturco). Indianapolis: Bobbs-Merrill.

Vico, G. (1988). *On the most ancient wisdom of the Italians: Unearthed from the origins of the Latin language* (trans: L. M. Palmer). Ithaca: Cornell University Press.

Wundt, W. (1916). *Elements of folk psychology: Outlines of a psychological history of the development of mankind*. New York: Macmillan.

Chapter 3
Historical Leads for Theory Construction in Psychology

Jaan Valsiner

Theoretical psychology needs to take the task of creating new theories seriously, and knowing history makes this possible. In the currently prevailing ethos of empiricism, history has become treated in psychology as if it is only a depository of lost and no longer useful traditions. The prevailing tradition—US undergraduate teaching-oriented habit of organizing history of the discipline into a "history and systems" classification of various "schools"—has organized our understanding of the history of the development of basic ideas in ways that make seeing their dialogical nature difficult (Valsiner 2012). The "systems"—labeled by some core notion (e.g. "behaviorism" or "cognitivism")—become employed as signs of mutually opposed "camps" in the fights for science. The continuity of theoretical concepts *across* the "systems" becomes secondary to the primary focus on the oppositions *between* them.

The heritage of Giambattista Vico (Tateo 2015) is a good example of how classification of key doers of psychology of the past leads the further development of key ideas astray. His "new science" can be hailed as the very root of the "sociocultural paradigm" in psychology, but knowing that does not allow us to understand the potential of his ideas for building indeed a "new psychology"on the basis of the *social* nature of human beings. Likewise, our understanding of agency in psychology has been limited by failing to consider a person as a whole—and opt for dividing the psyche into various mental or bodily separate states. Yet—as Smith (2015) emphasizes here—agency is the main characteristic of persons and needs to be studied as such.

3.1 History of Psychology is the Tool for the Future of Theory Building

What would be the forward-oriented role of history of psychology as a tool for development of the discipline? It is a tool of reflexivity—focused on the past, but oriented to the future. A careful analysis of a theme important in the past—yet

J. Valsiner (✉)
Niels Bohr Professor of Cultural Psychology, Department of Communication and Psychology, Aalborg University, Aalborg, Denmark
e-mail: jvalsiner@gmail.com

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_3

Fig. 3.1 Basic ideas for being a human being

abandoned by the epistemic markets due to their fluctuations—is to inform the future reconstruction of the discipline. For example, at our present time we face the need for reconstructing the introspective method as the core of human psychological research methodology. It had been eradicated from psychology under the attack of ideologies (the "behaviorist avalanche"), limits of the method itself (how to deal with the "imageless thought") and social macroprocesses of both world wars in the twentieth century and their corresponding refocusing of the social on socially massive phenomena—crowds in revolutions and wars, evaluation of persons within "mass ornaments" of armies, job candidates, or employees. Yet, it is needed again, over a century later—the focus on agency cannot be built without it. The study of agency needs a fresh methodological start.

3.2 Where Agency Begins

In analogy with the biological notion of "stem cells," we can posit the existence of the most basic ideas of the self-reflexivity of human beings: I→AM, I→WILL, I→NEED, I→WANT. In terms of their relations, we can set up the basic structure of their relationships (Fig. 3.1)

Psychology has historically concentrated on the I→AM part under the disguise THEY→ARE. The researcher has been likely to be left out of the study of the psyche, in favour of the study of some others ("THEY" = "subjects" or "research participants"—Bibace et al. 2009). The agency of the researcher him-/herself is traditionally downplayed—even if it is actually the central trigger of all data through the construction of any method of investigation. The goal of the researcher is to make sense of him-/herself (I→AM), yet the evidence through which such understanding can be reached is alien to him or her (plural: THEY→ARE is turned into generic singular SHE/HE→IS, yet maintaining the distance from I→AM). Agency cannot be successfully introduced into scientific psychology unless the sequence of helical transition {I→AM → THEY→ARE → I→AM} becomes exemplified. The researcher is the creator of knowledge through his or her own existential experiencing, even if the latter is for methodological purposes delegated to some others ("research participants"). The primary participant in the research is the researcher himself/herself.

3.3 Agency is a Temporally Located Concept

Even if the I→AM connection is specified, the focus on agency is still not complete. It takes the analysis of all parts of the structure in Fig. 3.1, with the highest level of the hierarchy (I→WILL) being crucial. Agency involves the active move towards the unknown future, based on one's needs and desires. Yet, each and every manifestation of agency is bound to the always unspecified PRESENT in the irreversibility of time that flows from PAST to FUTURE. Any moment of I→WILL cuts across that imaginary border of the PRESENT—precisely at the location of the connector (→):

> **I [→** the border of **PAST** and **FUTURE** which is the **PRESENT →] WILL**.

Thus, the question of agency acquires its presence in time, and therefore becomes the link of human beings with themselves in development. By adopting the theoretical centrality of agency, psychological theory construction becomes necessarily based on the general notion of inherent intentionality that was suggested by Franz Brentano. Yet, Brentano failed to see the severe theoretical constraints that time sets upon agency. Psychology, in general, needs to become developmental in its core so that it can accept the central role of agency.

3.4 What Renewed Focus on Agency Might Accomplish

A renewed focus on agency can perhaps bring psychology out of its current stalemate, well described by Steinar Kvale:

> Psychology entered the twentieth century as a promising young science, with new experimental laboratories being established and Freud's *Interpretation of Dreams* instigating a new psychological culture. At the start of the twenty-first century, however, the science of psychology appears in a puzzling state, somehow empty of radically new insights into the human situation (Kvale 2003, pp. 597–598).

The focus on agency gives psychology a chance, and theoretical innovation needs to precede empirical practices in this major reconstruction effort. History of psychology gives us a lead for it.

References

Bibace, R., Clegg, J., & Valsiner, J. (2009). What is in a name? Understanding the implications of participant terminology. *Integrative Psychological & Behavioral Science, 43*(1), 67–77.

Kvale, S. (2003). The church, the factory and the market: Scenarios for psychology in a postmodern age. *Theory & Psychology, 13*(5), 579–603.

Smith, R. (2015). Human agency and the history of psychology. In C. Gruber, M. Clark, H. Klempe, & J. Valsiner (Eds.), *Annals of theoretical psychology* (Vol. 12). *Constraints on agency: Explorations of theory in everyday life*. New York: Springer.

Tateo, L. (2015). The providence of associated minds: Agency in the thought of Giambattista Vico and the origins of social and cultural psychology. In C. Gruber, M. Clark, H. Klempe, & J. Valsiner (Eds.), *Annals of theoretical psychology* (Vol. 12). *Constraints on agency: Explorations of theory in everyday life*. New York: Springer.

Valsiner, J. (2012). *A guided science: History of psychology in the mirror of its making*. New Brunswick: Transaction.

Part II
Neurosciences look at Agency

Chapter 4
Neurobiological Perspectives on Agency: Ten Axioms and Ten Propositions

William R. Klemm

Understanding agency begins with the recognition that agency typically consists of multiple elements, such as the generation of intent, identification of contextual salience, recall of relevant memories, evaluation of choice options and their anticipated consequences, decision making, planning, implementing the action, and real-time monitoring and adjustment of the implementation. These elements of agency can be understood either from the perspective of neuroscience or philosophy. Traditionally, the most commonly described perspective seems to be that of philosophy, and there is abundant scholarly literature in that arena, which is reviewed in another chapter in this volume. The need for a more neuroscientific perspective seems evident. This chapter is an attempted response to that need.

The neurobiology of agency has typically focused on details of movement control: interactions of the basal ganglia, motor cortex, and cerebellum in generating and controlling movements. There is a conspicuous need for continuous updating of our understanding of such operations, but space limitations prevent such consideration here. In this overview, at the behest of the editor, I choose to explore axioms and propositions that might help explain the state of agency and guide future research. That is, I explore the state of agency in terms of what it is, how it is initiated, its neural correlates, and the differences between unconscious and conscious agency. Finally, I try to identify promising areas of research and suggest appropriate strategies and tactics.

4.1 Neural Basis of Agency

The concept of agency is fundamental to the existence of each individual operating in an environment that requires the individual to engage. Agency is not limited to consciousness or to humans, but occurs even in primitive animal species. It is, by

W. R. Klemm (✉)
Texas A&M University, College Station, TX, USA
e-mail: wklemm@cvm.tamu.edu

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_4

operational definition, applicable to whole-animal nervous systems. A spinal reflex, for example, might not be considered an expression of agency because it arises from local circuitry and is not necessarily engaged with the whole nervous system.

Colloquially, agency is considered in terms of what we do and fail to do, as I have summarized elsewhere (Klemm (2013). Yet, the topic is seldom discussed explicitly in scholarly works, and I have yet to find the word "agency" in any index of a textbook of neuroscience. Though the implicit meaning of the word is to act in the world, the behavioral expression of agency results from mechanisms that enable or promote it, as well as those that suppress it. Much of my own research has been devoted to brainstem-activating functions that enable agency (Klemm (1990) and to behavioral arrest functions that restrict agency (Klemm (2001). Now, I relish the opportunity, albeit with trepidation, to reflect on agency itself.

You could define agency as any animal action that arises out of its nervous system. It is instructive to consider at least in passing the evolutionary aspects of nervous systems. Consider simple jellyfish. They have no brain, but they do have a net of neurons that spontaneously discharge to propel the animal through seawater like an umbrella opening and closing. This kind of oscillatory agency occurs without stimulus or any self-organizing control other than the refractory period of impulse activity that is essentially simultaneous in all the neurons for a given phase of the rhythmic beat. Inherent pace-making neuronal discharge drives the species-specific rhythms.

Of all that I could say about such nerve nets, I emphasize oscillation here because it is a fundamental property of the brain, as we shall see later in this chapter. Intuitive assumptions about agency are based on cause and effect relationships such as "top-down" and "bottom up." Yet in networks, the nodes that share in reciprocal oscillatory drive reveal no clear cause–effect relationship.

The jellyfish nervous system has no capacity for intention or purpose, as we use the words. The whole point of this animal's agency is to move continuously, which has the effect of increasing access to algae and other microflora the animal can use as food. Such a creature exists to gather energy sources, capture the energy, and reproduce. Higher animals with true brains can create intentions, purposes, and make decisions that fulfill purpose. Human brains have a self-organizing activity that can create purpose consciously.

Agency arises from the processing reactions that take place at all levels of the nervous system. At each level, there can be substantial information modification. The most obvious information carriers are neuronal action potentials and the associated molecular interactions involved with neurotransmitters. The simplest form of agency processing at a circuit level is reflex action, where sensory input more or less induces an automated response, often with only a minimal amount of processing. When a reflex circuit is arranged so that the output goes to another reciprocally connected circuit, the basis is laid for reciprocal action, wherein the two circuits can interact, often in mutually regulating and even oscillatory ways.

All processing modification typically requires the activity of certain strategically placed neurons that can cause inhibition. Inhibitory mechanisms create the opportunity for inhibitory routing or gating. In reciprocally connected circuits that

have some inhibitory output, each reciprocal member can cyclically pace the activity of the other. Inhibitory routing also helps direct information flow selectively through certain circuits and not others, helping to ensure that there is parallel, multilevel processing.

When reciprocal action occurs in large ensembles of linked neuronal circuits, there is a functional basis for feedback and re-entrant mapping. Feedback occurs when part of a neuron's output is led back into that same neuron. Re-entrant mapping is a similar idea at the population level: Some of the output of one population of neurons projects through mapped pathways into a second ensemble, which after a certain processing delay, sends a re-entrant input into the first ensemble.

Parallel, multiple-level processing is most profound in the cell-dense cerebral cortex, where processing often involves segregation of sensory information into distinct but contiguous specific circuits called cortical columns. The combination of activity patterns in multiple pathways forms the basis for combinatorial pattern coding that is probably employed throughout the brain.

Activity in multiple pathways often exhibits rhythmicity and synchronicity because the widespread circuits have distributed circuits connected so they can influence each other. All of the foregoing processing phenomena and particularly re-entrant mapping, parallel distributed processing, and combinatorial pattern coding give rise to the many apparently emergent properties of the nervous system. We need not dwell on the details of these core operations to realize that collectively they provide the basis for agency.

Higher animals engage the environment both consciously and unconsciously. Engagement would seem to involve sequential active acts of generating intent of some kind, deciding which of the available action options to pursue (which entails value assessment and recall of relevant memories), planning how to implement the decision, and launching, monitoring, and adjusting implementation. In the case of conscious engagement, a person needs to pay attention to the requirements of each element. In short:

> Intend, remember, value, decide, prepare/plan, and act.

Each of these elements of agency is distinctly different. Yet, most agency-related research seems to reflect an investigator's focus on the observable action, as it if were a single point process.

Many decisions result from a preexisting intent to choose among available options and act accordingly. Other decisions can be triggered as a complex reflex by external compulsion without prior intent. The mechanisms of intent and decision are likely to be different and both likely differ from planning and execution of acts. Even decision making contains within its process the need to evaluate salience, expected outcome, reward/punishment value, relevant memories, and working memory. Planning and implementation of a decision require a different level of neural processing. Also, monitoring and adjustments in implementation are integral to willed action. As explored later in a proposition about free will, experimenters testing free will commonly ignored these elements of agency.

4.2 Axioms Versus Propositions

As with any topic being evaluated, it seems useful if not essential to begin with identifying what we think we know, what might be correct but as yet unconfirmed, and where new information and insights are needed. All three dimensions apply to the subject of agency. This analysis presents a series of ten axioms followed by a set of ten propositions, followed in turn by suggestions for neurobiological research. By definition, an axiom is a premise so evident as to be generally accepted as true without controversy. As used here, a proposition, on the other hand, is a declaration open to analysis and debate in the hope of eventually developing consensus on a derived axiom.

4.3 Agency Axioms

One should present axioms with trepidation. Some axioms stated here may well be subject to challenge, given that they had to be constructed without help from elsewhere. I do not know of any published listing that could serve as a guide.

Agency is a biological imperative. Some writers have viewed animals as inherently stimulus seeking. Certainly, that is true inasmuch as the biological necessity for food and reproduction requires animals to seek out stimuli that signal appropriate sources. Even animals, like ocean coral, thrust tentacles out of their exoskeleton in order to trap plankton as it drifts along. Evolutionary pressures typically favor species that are active in their world. Nervous systems evolved to expand the ability of species to be active agents in their environment.

In humans, these agency imperatives can be extended to include the notion of initiative, which seems to be an inherent property of the brain. Initiative is some kind of biological drive, existing in a widely varying extent in people, for which we know very little about its neurobiological origins. It is clear, however, that initiative drives human agency.

The axioms presented here are the author's opinion of the axioms related to animal and human agency, based on his 50 years of research and teaching of neuroscience, and his impression of what constitutes a consensus. One problem in arriving at such axioms is that the neuroscience literature rarely considers human agency as such, but rather the research is usually more focused on brain functions in other contexts. For example, movements have historically been classified as either voluntary or involuntary, and research has been conducted on the movement control mechanisms in that context rather than in the more general case of human agency.

A1. Agency Arises When a Nervous System Has a Sense of Self, Which May or Not be Conscious. Humans Have a Dual Sense of Self, Unconscious and Conscious

The nervous system, even in primitive organisms, has a built-in anatomical mapping of where its various body parts are in space and time. They even have sensory systems that in the process of detecting biological relevant aspects of the environment also inform the nervous system of the self and nonself.

In humans, the sense of self (SoS) is like a sixth sense functioning with the same neuroscience principles used by the other senses, yet distinct in that it individuates our other senses in complex and rich ways (Klemm 2012). SoS is actually created by the other senses. All six senses begin early during nonconscious development in the womb as a result of embryonic cell division, migration, and differentiation of neurons. Sensory and motor topographical maps are formed. The sculpting of early circuitry is influenced by self-referential stimuli from the developing fetus and from mother-specific signals from the womb that inform those circuits that they have a body they can influence. At some point, about 2 years after birth, enough neurons appear and circuitry differentiates to enable episodes of conscious awareness of self and nonself (Povinelli and Giambrone 2001).

Consciousness occurs in the context of self and nonself-awareness. I think that much of this self-consciousness is implicitly learned. Such learning, to be consciously operative, requires a species to have enough neuronal circuitry to create the information "carrying capacity" required for conscious representation of ordinary stimuli, an SoS, and cognitive interactions with environment. Such a capacity arises developmentally.

The neurobiological evidence that each individual sculpts its selfhood and capacity for agency provides clear support for existential philosophy, which holds that existence precedes essence. A human first comes into existence, as a fetus and then baby, but then becomes what experience makes the person to be. Agency need not be contemporary, but often emerges from memory stores that are recalled by a nervous system for current application.

A2. Agency In Primitive Animals Is Caused By Central Pattern Generators, Remnants of Which Remain In Higher Animals

Sometimes, behavior is driven by sparse "command neurons" such as Mauthner cells in the fish brainstem that automate tail swishing. Such acts of agency arise from how the neural circuit is anatomically constructed. For complex behaviors, circuit-level decision making must occur but may ultimately be expressed via central pattern generators. While little is known about circuit-level functions, elegant studies with a dissected, but functionally intact nervous system of the medicinal leech have disclosed the existence of decision-making circuits (Briggman et al. 2006).

In a spinal ganglion of the leech, investigators stimulated the inputs that drove mutually exclusive crawling or swimming behaviors. Monitoring large populations of neurons with voltage-sensitive dyes revealed single neurons that discriminated the stimulus in advance of a decision to crawl or swim. One neuron was found that biased the leech to swim when the cell was hyperpolarized and to crawl or delay swimming when it was depolarized. Although this neuron has been called a command neuron, these studies made it clear that it is part of a decision-making circuit that determines which of the two behaviors will occur.

A3. Agency Emerges From Neural Processing

Processing reactions in higher animals take place at all levels of the nervous system, from the coding in sensory receptors, to synaptic reactions and associated patterns of impulse generation, to the routing through neuronal networks, to profuse interactions of widespread neuronal subsystems.

Bottom-up drivers of agency involve sensory input that is sufficiently salient to require a behavioral response. Top-down agency is typically thought to arise in the cerebral cortex, as processing of reinforcement contingencies, memory, action planning, and the like coalesce to generate behavior. Much of top-down processing has been ascribed to synchronization of activity within multiple cortical areas (Von Stein et al. 2000).

Such processing typically involves coordinated activity within specific regions of the brain. Numerous examples of how various brain regions participate in initiating and directing action can be found in any modern neuroscience textbook.

Consciously directed agency, often referred to as executive control, is presumed to require participation of restricted domains of cerebral cortex. Obvious examples include the well-known cortical “centers,” such as speech centers, face recognition areas, and motor-cortex topographical maps.

One recent study of interest has focused on the role of memory, in the context that actions often depend on what one remembers (Paz-Alonso et al. 2013). This was a functional magnetic resonance imaging (fMRI) study of 43 children and young adults who learned 56 word pairs and were then shown one of the words and asked either to recall the associated word or to prevent the associated word from coming to mind. Regardless of age, individuals who were best able to suppress memory recall exhibited tighter activity in coupling between the hippocampus and several cortical areas (dorsolateral prefrontal, cingulate, and parietal). The ability to suppress recall seemed to improve with age. It is not clear whether such ability progresses automatically with neural development or is learned.

A4. The Currency of Information Processing in the Brain is the Pattern and Timing of Nerve Impulses (Circuit Impulse Patterns). Therefore, the Elements of Agency Must Arise From Biological Computations That Change Such Patterns

If thinking is the flow of CIPs (Klemm 2011a, 2011b), how does that enable us to make intentions, choices, or decisions? As a representation of the current mental state, circuit impulse patterns (CIPs) can be changed by external input or by feedback as the CIP codes are routed through various sub-circuits and modified in the process. Let us remember what intentions, choices, or decisions really are. They too are CIP representation that can be propagated to generate new thought, intent, or decision.

The particular spatiotemporal patterns of CIPs are initially generated by stimuli (or memory of prior stimuli). Higher animals can modify CIPs in ensembles of neurons because of the rich capacities of synaptic physiology involving transmitter release and postsynaptic receptor binding. Whether these ensembles act as control "centers" or not, they never operate in isolation. Any hierarchical operation is subject to re-prioritization.

A5. Agency Arises Out of the Multiple Nonlinear Deterministic Processes of the Nervous System

Neuroscientists have established that the nervous system operates via nonlinear deterministic processes. "Deterministic" means measurable cause and effect, which in theory could be fully described by equation. However, the brain's nonlinear determinism is so complex that there is little hope of developing equations, attempts within chaos theory notwithstanding. This brings us to the heart of agency and raises intriguing questions about "free will" (see Proposition #8).

A6. Behavior is Not Just the Response to Stimuli. Though the Most Obvious Face of Agency, Behavior is the End Product of the Antecedent Elements of Agency

Organisms exist through their interactions with the environment, and thus stimulus–response relations alone are regarded as inappropriate units of analysis. Not all behavior is directly linked to an immediate stimulus, and even when so linked, agency is not a point process but rather smeared out over time to allow processing of the necessary elements of agency.

The behavioral response to activation of a given sensory receptive field is variable. That is, the response to stimulation depends on the behavioral context in which

a stimulus occurs. Such variability has been amply demonstrated in many kinds of receptive fields. The nervous system can adjust behavior, as in orienting, to maximize detection of stimulus. The brain has ample ability to adjust both central and peripheral neurons to the demands of the behavioral context. Thus, all neural activity is active—not just reactive.

Reactivity, however, is a fundamental building block of nervous system function, as has been amply demonstrated in many experiments in which portions of the nervous system have been isolated, as in spinal cord transection. Just because stimulus–response paradigms are simple does not mean that they are complete. Animal brains have multiple functional systems to process that information as the dynamical mechanism for generating appropriate and successful behavior.

A holistic, systems-level perspective provides just another way to say that the whole is greater than the sum of its parts. We should have no problem with believing this conclusion. Nonetheless, the whole is not independent of its parts, and an understanding of its parts may prove to be a prerequisite for understanding the whole.

A7. Neural Correlates of Agency are Necessary But Not Sufficient for Explaining the Various Elements of Agency

Some four decades ago, when I entered the scholarly fray over the significance of hippocampal theta rhythm, the prevailing dogma was that theta was a specific correlate of voluntary motor activity. Thus, it was perceived as a cardinal index, and perhaps cause, of this aspect of agency.

I and others showed that theta sometimes correlated with involuntary motor activity, and under animal "hypnosis" conditions, even immobility (Klemm 1976). We now know that theta correlates with many neural functions and has a much more fundamental and far-reaching significance, some of which is discussed later in this chapter.

The obvious point that I made then that applies here is that correlation is not causation. The cause of any given element of agency (intent, valence, etc.) has its correlates but not all of those correlates will be the cause of the agency element.

A8. Initiating Agency, Such as Deciding What to Do, Typically Requires Assessment of Anticipated Positive And Negative Consequences And Their Reward Value

Humans tend to assign value to different options for a given agency possibility. While abundant psychological and economic research is relevant, we know little about how the uncertainty of reward or risk is represented in the brain and influences decision making. Nor do we understand much about how the brain handles the

trade-off between the overall magnitude of the reward and how immediately one receives the reward.

Nonetheless, rewards are often the major parameter that affects decisions. Reward could be immediate or delayed. Neuroscientists want to understand how neurons represent rewards, and how information on rewards is integrated over time in order to reach a decision. One thing seems likely: Multiple areas of the brain participate in integrating positive reinforcement expectations in self-initiated behavior. This was recently demonstrated in studies involving roulette gambling (Studer et al. 2012). In that case, fMRI imaging implicated the midbrain, striatum, anterior insula, ventromedial prefrontal cortex, and inferior parietal cortex. Though this particular study involved conscious decision making, similar distributed processes are likely involved in determination of unconscious agency.

Another recent fMRI study evaluated what happens in the brain when music gains reward value the first time it is heard (Salimpoor et al. 2013). The best predictor of the amount of money people will spend for music in an auction paradigm was increased activity in the nucleus accumbens, a well-known component of the famous medial forebrain bundle "reward center." When listening under conditions involving evaluation, activity also increased in the amygdala, auditory cortices, and prefrontal regions, though this activity did not predict reward value. However, such a value was predicted by the level of connectivity of these areas with the accumbens as reward value increased.

A similar approach could be developed to illustrate how agency can be driven by the inherent biological aversion to negatively reinforcing contingencies.

A9. Consciousness, Especially Human Consciousness, Raises the Level of Agency Complexity to a New Level

Now, not only are basic neural circuit functions in play, but issues also arise about will, and especially the degree, if any, to which acts of will are freely chosen. Another issue is the common assumption that consciousness is more than a mere observer. A growing fad in neuroscience is a belief that consciousness cannot do anything. Elsewhere, I argue vigorously against such a position (Klemm 2014).

A10. The Human Brain Contains a Distinct Network that Serves as its Executive Agent

This network is primarily based in the dorsolateral prefrontal, parietal, and cingulate cortices. It regulates the many "top down" neurobehavioral functions that are so characteristic of the human brain (Banich et al. 2009). Deficiencies in the function of this network underlie numerous neuropsychiatric conditions (Beck 2008). The ability to regulate emotions and direct rational actions is typically associated with success in life, and an inability to do so often leads to dire consequences.

This network can be trained to develop a more robust capacity for executive control. This, as we all experience, is what parenting and schooling are about. Such training is especially crucial in early childhood as the challenges of school are first encountered. Even so, such training takes many years and for most of us may never be completed.

The question arises: Can such executive control training be expedited? One possibility has recently arisen from several studies showing that working memory capacity can be expanded by a relatively short training time, and in the process general intelligence may be improved. Since the same system that determines intelligence is also operative in executive control, Schweizer et al. (2013) reasoned that working memory training might also enhance executive control. To pursue this possibility in a specific context, researchers hypothesized that inappropriate or maladaptive behaviors might be reduced by effective working memory training based on emotion-laden stimuli.

In this study, subjects in their early 20s were assessed for affective control before and after 20 training days of 20–30 min sessions. The experimental groups received dual *n*-back training with a simultaneously presented face and a word that was either emotionally negative or neutral. After each picture–word pair, subjects were to press a button to indicate if either or both members of the pair matched the stimulus presented *n*-positions back. Tests began with $n = 1$ and increased as subjects gained proficiency.

Not surprisingly, errors in both trained and untrained subjects increased at levels beyond $n = 1$, and the error rate was comparable for both groups. Results indicated that subjects reported less distress when they consciously willed to suppress it compared with the null state of just attending to negative stimuli. But this distress reduction occurred only in the emotional working memory training group.

No change in activity levels was indicated in fMRI scans as a result of placebo training, but significant increases occurred as a result of emotional working memory training irrespective of the level of *n*-back achievement in the executive control regions of interest.

The study also compared emotional responsivity before and after training. Subjects were asked to just pay attention or to pay attention and cognitively suppress their emotional reaction. Subjects rated their emotions on a numerical scale from negative to positive while viewing films that were emotionally neutral (such as weather forecasts) or that were emotionally disturbing (such as war scenes, accidents, etc.). Training caused no change in the group that viewed only neutral images, but in the groups viewing disturbing scenes, training decreased the perceived distress in a group told just to attend to the scenes and was even more effective in the group told to suppress emotional reaction.

The affective working memory training produced benefits that transferred to the emotional response task. Trained subjects not only generated enhanced emotional regulation but also developed greater fMRI activity during the emotional task in the predicted brain regions of interest, the executive control loci. It seems that working memory training can do more than just expand the amount of information that can

be held in working memory. Emotional working memory training improves the ability to suppress disturbing emotional responses and does so presumably because the executive control network is more activated. Thus, such training might also enhance many executive control functions, particularly responses to emotionally disturbing circumstances.

4.4 Propositions

The following agency propositions come from ideas currently in vogue among neuroscientists. All of these have a degree of controversy, and some (#7 and #8) are not endorsed by this author.

P1. The Correlates and Causes of All Elements of Agency are Discoverable and Can be Found in CIPs

The inputs received by the brain from sense organs are abstracted and represented by patterns of nerve impulses flowing in specific neural circuits. This representation can be stored via synaptic changes in neural circuitry for later "online" replay in the form of memory recall. All the elements of agency must use this impulse language, and each element may be represented by corresponding CIPs.

We know that movement occurs because of the impulse patterns in antecedent motor preparation, whether conscious or unconscious. This preparation takes into account the existing state of muscle tone, the existing position of limbs and joints, and the required (or desired) future position of the limbs and joints. All such information is represented in CIPs. The motor component of agency likewise is caused by nerve impulses. Motor commands are ultimately expressed through a final common path that terminates in the neurons that make direct contact with muscle or glands.

Willful intent to move is self-evident to humans. And similar intent has been clearly demonstrated in a classic experiment in animals as well (Evarts and Tanji 1976). They trained monkeys to move a lever in a certain direction when they received two successive cues. The first cue told the monkey the direction to move the lever, and thus served as a signal to the brain to "plan" a future motor act. After a certain delay, the second cue told the monkey to produce the movement. Extracellular recordings from neurons in the motor cortex showed that some of those cells became active after the first cue, before the second cue was presented. These neurons are thought to be generating the motor-planning program that is triggered into action by other neurons ("go" neurons) that respond to the second cue. Similar indexes of motor intent have since been reported in humans from strategically placed electroencephalogram (EEG) electrodes or by scans of cerebral blood flow.

This study, pioneering though it was, could create the misleading impression that intent is generated in the motor cortex as an inseparable aspect of agency. This is not likely the case (see Proposition #8).

P2. Decisions to Act Are Not Point Events But a Process of Multiple Stages. The Motor Act Itself Is Just the End Product of Antecedent Processing Elsewhere

The motor cortex is not, by itself, the seat of motor will but rather an executive agent for movement. We do not know how the various motor centers in the brain are orchestrated in developing motor will, but several studies in the current year show that these matters are amenable to experimental clarification.

Certain neurons in the brainstem can drive lower motor neurons in the spinal cord or motor neurons in cranial nerves to cause muscle contractions that subserve postural tone and body movements. That is, when higher neural circuits "make a decision" that certain body movements should be made, preprogrammed instructions help create a specified pattern of muscle contractions. In short, willed movements are created against the backdrop of automated adjustments in postural tone.

It appears that researchers have an increasing interest in exploiting discoveries that willed agency involves motor preparation, which includes prior decision making and planning that is not processed in the motor cortex. It is increasingly clear that important elements of agency precede actual movement and are reflected in neural activity signatures in various brain areas.

A recent fMRI study elucidates how humans make decisions regarding future implementing movements (Filimon et al. 2013). Eye and hand-movement preparation was seen to be separately processed in terms of perceptual decisions and the actual movement. Sensorimotor cortical areas were not involved in accumulating sensory evidence used in decision making. Rather, the inferior frontal cortex became conspicuously more active in the early decision-making stage. Moreover, once planning was under way, the intraparietal sulcus area joined in the motor preparation, but differently in different parts of the sulcus. Thus, it seems likely that different electrical signatures of motor decisions and planning occur in multiple brain areas before the primary motor cortex launches the intended action.

Conscious object-directed hand actions require action planning in other brain areas to integrate the goal with the required hand action. Such planning-related signals have recently been disclosed in human fMRI studies (Gallivan et al. 2013). Those experiments distinguished neural activity of multiple cortical sites during human hand movement (reaching for an object or actually grabbing it) under conditions of baseline, intention, and movement condition. Eleven regions of interest in each hemisphere established that activity during the intention and planning stage could predict which hand would be used and whether the action would be reaching or grabbing. Some frontoparietal areas in the planning state predicted which limb would be used, while other areas predicted the kind of hand action to be used,

but most of the areas coded for both kinds of prediction. Moreover, a substantial number of areas coded for ipsilateral movements as well as contralateral ones. Ipsilateral coding obviously depended on interhemispheric communication and could have represented a copy of the activity from the contralateral hemisphere, which in humans controls limb movement. Alternatively, ipsilateral activity might reflect parallel contingency planning.

Regardless, the main conclusion is that many cortical areas in both hemispheres develop intention—and planning-specific activity prior to specific goal-directed movements. The existence of such activity in both hemispheres provides a way for the planning to benefit from re-entrant processing.

Predicted outcome is also processed outside the motor cortex. A recent study of single-neuron recordings in monkeys distinguished choice coding that was guided either by movement outcomes predicted by stimuli and those guided by predictions of the animal's own actions (Luk and Wallis 2013). Monkeys were rewarded with tasteful juice either in association with a visual stimulus or with the hand movement of a lever. Neurons in both the orbitofrontal cortex and the anterior cingulate gyrus were capable of encoding the chosen action, but the orbitofrontal cortex was more likely to code when the choice was driven by stimulus, whereas the anterior cingulate was more likely to code when the choice is guided by the monkey's own actions.

P3. Decisions Are Made by Interacting And Competing Neuronal Populations or By Guided Gating Processes

It would seem that executive control over a response operates at two levels, perception of the stimulus situation and selection of a suitable response. Both processes can operate on different sets of information, and whenever the information is conflicting, the respective neural networks must resolve the discord in order for appropriate action to occur.

Alexander Soutscheck and colleagues (2013) recently explored the implications of earlier neuroimaging studies that revealed a correlation of increased neural activity under information conflict conditions. Those studies had shown that stimulus perception was associated with increased activity in the superior/middle frontal and parietal cortex (PPC), while response selection was associated with increased activity in medial and inferior frontal cortex. Thus, stimulus perception and actionable decisions are likely to be different processes.

To test whether these observations were merely correlations or causally involved in active resolution of the conflict, they applied transcranial magnetic stimulation (TMS) over each respective area to see if conflict processing could be disrupted. Magnetic stimulation of PPC selectively disrupted processing of stimulus conflict, while response selection was disrupted by magnetic stimulation of the pre-supplementary motor cortex (which others had suggested exerts a top-down control over motor cortex to resolve response selection conflict).

The idea of separate and dissociable decision networks does not exclude the possibility that certain decisions may be made by competitive, winner-take-all, processing in neural networks that commonly overlap. One of the most obvious ways that information is modified is by algebraic summation of inputs. If inputs have overlapping projections to common targets, then the effect on one input pathway can be augmented or suppressed by simultaneous input from another input pathway. The advantage of shared circuitry is obvious: Fewer neurons are needed to produce a range of behaviors. With a fixed number of neurons, overlapping circuitry allows more diverse processing.

One example of overlapping circuitry occurs in the abdominal ganglion of the mollusk, *Aplysia*. More than 90% of the neurons are active during reflex withdrawal of the gill. The same neurons are also active during respiratory pumping and during small spontaneous gill contractions. Selective behavior can occur because the temporal pattern of activity is different for each of the three behaviors. If all populations were active at the same time, occlusion would cause a less-than-optimal response. In other words, shifting timing relationships of activity in different circuits determines the specific nature of caused effects (Carr 1993).

Another consequence of overlapping circuitry is that the respective overlapped populations of neurons interact with each other. If the impulse activity in each overlapped populations represents alternative output probabilities, then the overlapping sets up competition among the various populations. As activity in each population intensifies, the competition inevitably favors one population (i.e. alternative outcome) at the expense of others. At some point, a threshold is reached in the favored population that constitutes a decision or choice of one among several alternatives.

A common view of how neural circuits make decisions is that different options are processed in separate but interconnected circuits and a resulting choice arises through competition (Fig. 4.1; reviewed by McMains and Kastner 2011). Most of the evidence for such a view comes from studies of visual perception.

Decision making has to begin from some kind of starting point after an intention to act is made. Based on computational models, it would appear that evidence accumulates in decision-making circuitry from the starting point to the ultimate decision.

A process that accumulates and averages noisy signals as new evidence is accrued has been invoked as central to decision making in neural populations (Brunton et al. 2013). In both rats and humans, an analytical drift-diffusion model enabled a moment-by-moment prediction of the temporal evolution of mental accumulator processing of noisy sensory signals. The model analysis suggests that behavioral decisions arise from noise reduction in the sensory processing, not in the evidence accumulator.

Lange et al. (2013) reasoned that prior expectations of possible outcomes would alter the starting point as well as the evidence accumulation process. To test the idea, they recorded magnetoencephalograms while inducing changes to the prior expectation. They observed that choice-selective motor cortex activity in the 8–30 Hz band was biased by prior expectations, along with biases in perceptual judgment. They postulated that expectation can alter the start point of the decision process,

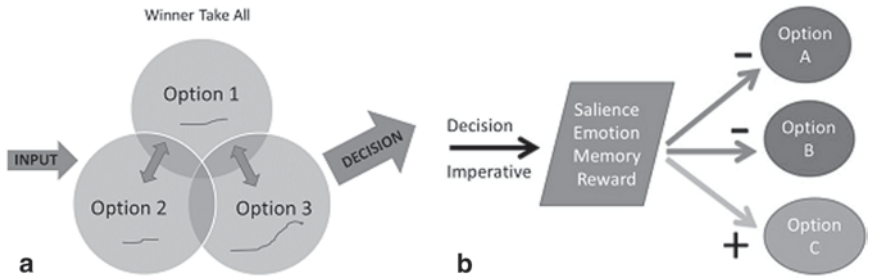

Fig. 4.1 *Left*: **(a)** Interacting neuronal populations could compete to reach a decision. Each alternative choice is presumed to be processed in its own local circuitry. The overlap with circuits processing alternative choices enables sharing of impulse traffic as each population competes for dominance. Activity within a population either builds up or diminishes as interaction progresses. At some point, a threshold is reached within one of the populations (*Option 3*), making its activity dominant and representative of the final decision. *Right*: **(b)** This is an illustration for an alternate guided gating approach to decision making. The stimuli that create the need for a decision drive evaluation of alternative options in terms of their salience, emotional context, memory of relevant past learning, and the expected reinforcement associated with each outcome. The evaluation steps may occur in multiple circuits, but they collectively guide the decision-imperative drive to one of the alternative options at the expense of others.

change the gain or rate of accumulation toward the expected outcome, or selectively lower the amount of evidence required for the expected outcome. Their particular study implicated a change in the start point.

A common explanation of the decision-making process is that various options are processed in separate local networks that are in communication with each other. Sensory stimuli and stored memories modify the activity in each local network, increasing or decreasing it. The interacting local networks may mutually regulate, so that as activity in one network builds up, it suppresses activity in the others. By whatever means, at some point the activity level in one network reaches a threshold where its output prevails in a winner-take-all fashion to yield a choice or decision (see Fig. 4.1, left).

An alternative to the winner-take-all view of decision making is the "guide-gating" or "global guidance" mechanism, in which lateral inhibitory gating guides the neural processing in the cortex as it progress to a final choice selection. The thalamic nucleus reticularis would seem to be central to such guidance (Taylor and Alavi 2003). Factors that influence the gating or bias toward a given option would certainly include biological relevance, emotional state, memories, and reinforcement contingencies (see Fig. 4.1, right).

One way these option-specific populations could be developing growing dominance among competing populations is an integrate-and-fire attractor network process. Computer simulation shows that such a mechanism is feasible (Rolls 2012).

Ancillary influences such as salience, etc. influence option selections in a major way. Salience determination has been attributed to a network involving the dorsal anterior cingulate cortex and bilateral insulae. In a study of how the network nodes

interact, dynamic causal modeling of human fMRI data indicated a way to disentangle the sequence of events in a network's components when stimuli co-activate the network nodes. One particular study, for example, demonstrated that the right anterior insulae has the central role in the network's response to errors, while the anterior cingulate cortex guides implementation and moment-to-moment adjustments of cognitive control (Ham et al. 2013).

Also, consider the dopaminergic reward system, for example. Repeated social stress can suppress generalized voluntary agency, creating social aversion. The effect has recently been found in mice that were stressed by aggression (Barik et al. 2013) or social isolation (Niwa et al. 2013). In both studies, the cause came from the release of stress hormones (glucocorticoids) acting on postsynaptic neurons that are responsive to the positive reinforcement actions of dopamine.

P4. Intentions and Decisions Require Some Mechanism for Orchestrating the Interaction Among Neuronal Populations, and a Likely Mechanism is Synchronized Oscillating Neural Activity in the Respective Regions

The extracellular currents associated with CIPs produce volume-conducted field potentials (at the scalp, this is called the EEG) and these in turn are large enough in many areas of the brain to influence the genesis and propagation of nerve impulses. Because of the underlying recurrent property of certain CIPs, oscillation arises from circuit self-organization, and such oscillation creates a mechanism for regulating neuronal activity. Recent evidence indicates that oscillation is a factor in the genesis of both the intent to act and of action selection.

Intent to act arises from needs of the organism and environmental contingency. For a brain to decide how to fulfill its intent requires evaluation of the salience and relative value of alternative choice options—all in the context of memories of past decisions. Because value assessment and memory are represented and processed in different brain regions (prefrontal cortex and medial temporal lobe respectively), and because both areas generate oscillating neural activity, a neurobiological mechanism for inter-area coordination could be achieved through synchronized oscillations, as manifest in coherent extracellular field potentials (Buzsáki and Draguhun 2004).

This idea is supported by a recent magnetoencephalographic study in 20 normal humans (Guitart-Masip et al. 2013). These researchers used a decision-making task devoid of spatial learning (to avoid confounding with spatial functions of the hippocampus) that included active-choice trials with different levels of reward and punishment. Comparison was made with a control task in which a comparable decision was forced. They showed that theta rhythm (4–7/s oscillations) in the hippocampus and three sites in the prefrontal cortex became synchronized specifically during decision making per se, independent of the expected value of any likely reward. No such synchrony was evident in the forced-action controls. Thus, at least in these

conditions, theta oscillation synchrony correlated specifically with the process of decision making. Since theta rhythms are extracellular reflections of underlying CIPs, the real process of inter-area coordination may be accomplished by shifting the time relationship of impulse activity in the respective areas.

Changes in the kind of action is typically associated with changes in field potential oscillatory magnitude and frequency. This has been known for cortical alpha rhythms for many decades. A recent study of human alpha activity revealed the staged participation of two executive control networks in the dorsolateral and dorsomedial parietofrontal pathways (Verhagen et al. 2013). They monitored EEG recordings from scalp regions over the corresponding pathways and interference of pathway function by TMS under conditions wherein grasping an object was made relatively difficult or easy. Immediately after a signal to initiate grasping, the EEG was recorded and a low-intensity single TMS pulse was applied to one or the other pathway scalp region (or a control site at the vertex) during two time windows.

Typically, EEG alpha oscillations decrease as processing demands increase. This suppression was more pronounced over the medial pathway with the more difficult grasping task during the first second after the signal to start movement. The suppression then persisted for the first half of the movement. TMS over either pathway disrupted the enhanced EEG activity along the medial pathway: early when the lateral path was perturbed and later when the medial path was perturbed. Thus, even though both paths participated in processing the movement, the dorsomedial pathway seems to depend on the computational operations in the dorsolateral path. TMS over the medial path enhanced alpha in the lateral path in the planning stage, suggesting that the lateral path can compensate for diminished processing in the medial pathway. One interpretation is that the lateral path first constructs the plan for grasping and the medial path guides the execution. If the medial path loses functional capacity, as from TMS, the lateral pathway can compensate the movement guidance.

The degree and time course of alpha suppression was also changed by TMS of focal cortical areas that control goal-directed movements. The TMS manipulations were designed to dissect hierarchical relationships among two parietofrontal networks that share the control of goal-directed action. Either network can participate in the reach and grip components of grasping, but they may operate at different hierarchical levels. Single-pulse TMS over either network disrupted the alpha suppression that was triggered by increasing the grasping task difficulty but at different times, appearing first within 1 second in the dorsomedial circuit (when movement was being planned), then in the dorsolateral circuit (when movement was under way).

What drives neural networks into oscillation? We know that networks with architecture that can support oscillation can be driven by outside input or can emerge from within a self-organized network system. These processes are mediated at the neuronal ensemble level, not the level of single neurons.

Unknown ancillary factors can contribute to oscillatory properties in neural networks. There is a recent report that injecting dopamine into prefrontal cortex of anesthetized rats induces synchronization with an ongoing hippocampal theta

rhythm (Benchenane et al. 2010). This has no clear relation to agency, because anesthesia precludes agency. However, such a condition reminds us of the utility of comparing what we learn about mechanisms of agency with brain functions during agency-blocked states like sleep and anesthesia.

P5. The Unconscious Mind is a Brain Agent

Primitive animals whose nervous system cannot support consciousness nonetheless are active agents. I recall recently watching a train of foraging ants, moving in two opposite trains along adjacent parallel paths. One train was carrying food to the nest, the other was returning from dumping their food load to go get more. Most amazing to me was that ants leaving the nest often bumped into inbound traffic and paused a fraction of a second to identify and apparently recognize ants bearing food. Never did the outward-bound ants try to steal the food, as if they understood they still had an unfulfilled task to perform.

Such action exemplifies primitive agency. At the most basic level, reflex action is a relatively simple involuntary and stereotyped response to specific stimulation. A reflex action results from relatively invariant connectivity between the sensory input neurons and the motor output neurons.

Many "higher" nervous system functions are based on the integrated interaction of several or more basic reflexes, and they may also be involuntary and stereotyped. These compounded reflexes may be summed or even occur as a series of linked or chained reflexes, which many people prefer to describe as "reactions" or "responses."

Such responses are commonly called fixed-action patterns, which are found in the behavioral repertoire of many animal species. These patterns are generated by central pattern-generator circuitry in response to rather specific stimuli or environmental contingencies. These also occur in higher animals, particularly affecting mating behavior (Rivard and Klemm 1990). Fixed-action patterns may subserve more complex (and less reflexive) patterns of motor activity, such as human habits.

P6. The Conscious Mind is the Brain's Observer of Some Portion of Unconscious Mind Operations

The "observer" nature of consciousness is central to the argument that the role of consciousness is to create an awareness of the consequences of unconscious agency (Gazzaniga 1998). This is reminiscent of the Cartesian Theatre theme where the consciousness allows the brain to see some of what is happening on its stage. Contemporary scholars commonly modify the original Theatre concept by positing that consciousness involves some kind of inner agent which does the viewing. I view consciousness as a unique set of CIPs created by the brain as an avatar that can act on

behalf of embodied brain (Klemm 2011a, 2011b). However, many scholars argue that the conscious mind is only an observer and cannot do anything (see Proposition #7).

One perspective I do not see discussed much in the neuroscience literature is the fact that unconscious and conscious minds are dramatically different yet are seamlessly entangled. Theorists may have boxed themselves in by the semantics of the conscious and unconscious. Maybe there is only one mind, operating in shifting state changes to which we have attached labels. It seems that conscious operations are explicitly deliberative, analytical, focused on the novel and unlearned, and relatively slow. Unconscious operations seem to operate at high speed on well-learned memories.

Then there is the matter of wakefulness, which in humans is typically associated with consciousness. Consciousness does not occur in sleep (except in dreaming, which is not really sleep, and in fact may occur as the brain's way to wake itself up and "reboot" consciousness—Klemm 2011c). The unconscious mind operates in sleep, as verified by the numerous experiments showing that memory consolidation is facilitated during sleep. Wakefulness is a state where both unconscious and conscious processes operate at the same time. If the agency of wakefulness is not accompanied by conscious thought and directed action, then we humans are either zombies or robots. The usual definition of zombies is human existence without consciousness. Since we know, at least for ourselves, that we are sometimes conscious, this means we are biological robots, assuming that conscious thinking cannot cause anything.

P7. Conscious Mind Is Not a Source of Agency

This proposition is framed in an apparently untestable way. How does one prove a negative, especially this kind of negative? Nonetheless, prominent scientists argue that consciousness is worthless and does nothing (see Haggard et al. 2002; Wegner 2005). Such scientists concede that neural events cause agency and consciousness, but consciousness cannot cause neural events or agency (Pockett et al. 2006). But what if consciousness, as I suggest, is itself neural events?

Those who believe consciousness cannot cause anything hold that consciousness is an epiphenomenon of brain function and as such can only make explicit some otherwise unconscious thought. I have found no actual evidence for such a view, other than speculative extrapolation from simple reflex-like movements purported to disprove the idea of free will (see Proposition #8). But even if one accepts the position that free will is illusory, how does that inevitably support the position that there be no conscious will at all?

One obvious objection to this proposition is that brain and body properties typically evolve because they satisfy some natural selection advantage. Is there no advantage to consciousness? Which is more adaptive: to evolve capacity for false belief in our powers of introspection and consciously directed agency or to actually have such powers? Would we be just as functional as zombies?

I also raise this question: Why is there no agency when we are asleep? If the conscious mind cannot do anything, why does agency only occur when we are conscious? The contrived answer might be that unconscious mind cannot do anything because it is asleep. If we could wake up without consciousness, functioning like zombies, then it would prove that consciousness is not needed to cause behavior. Of course, all awake humans act like zombies on occasion, but for most people that state does not dominate wakefulness.

Common experience teaches that agency does not occur in unconscious states such as sleep, coma, or anesthesia. If agency only occurs during consciousness, how can one assert with any assurance that consciousness is not relevant to agency?

What consciousness does, if anything, has to relate to two key aspects of consciousness: making thought explicit and reducing information density in low-capacity working memory.

Surely, consciousness might through its explicit restructuring of its worldly representations provide another way for the unconscious mind to be taught and programmed. I should think that executive control is facilitated by the explicit awareness, introspection, and analysis afforded by consciousness. Yet this rather obvious conclusion has been recently challenged (see overview of special topic papers by Di Pisapia, 2013).

The possible agency-related functions of consciousness include: (1) serving as a scratch pad for selection of choices and agency, (2) long-range planning, (3) construction and storage of memories, (4) retrieval of stored unconscious memories for use and modification, and (5) "troubleshooting" and reflective analysis and decision making (Mandler 2003).

Consider the role of consciousness in learning and memory. Common experience establishes that focused attention, conscious rehearsal, and use of mnemonic devices promote memory consolidation. True, many memories if sufficiently rehearsed in consciousness become stored as unconscious procedural memories, as evident in the formation of habits, prejudices, and motor memories like touch typing, riding a bicycle, and sports skills. An established axiom from memory research is that information needs to be learned in small chunks because working memory capacity is limited. Information that is not consciously rehearsed is less likely to be remembered. Eventually, with sufficient conscious rehearsal, the learned skills become incorporated into unconscious mind such as habits, prejudices, and motor skills.

This brings up the point of the reflection, information organization, analysis, planning, and the like that occur during consciousness. Do these processes not impact what is going on unconsciously? It seems natural to suggest that the main function of consciousness is to refine, amplify, and perfect unconscious processes. Yet, the original premise is crucial. If consciousness is only a display mechanism, then of course it could not do any of the things we usually attribute to it.

When we use consciousness's capacity of introspection and conscious analysis, we increase the odds of spotting flawed logic and increase the options for better alternatives. Unfortunately, the unconscious mind's natural tendency is to believe its initial conclusions and not to second-guess them. In short, a main value of consciousness is that it enhances the quality of feedback for unconscious processes.

One can also make a neurophysiological claim for consciously directed agency. Gray (2004, pp. 114–117) summarizes three experiments that support the notion of causal consciousness. Yet he admits that these studies are not conclusive.

Freeman (2000) contends that consciousness "must play a role in intentional behavior," because consciousness is a neural process that embeds patterns of impulse and field potential representations of past and present events and states. In his perspective of chaos theory, Freeman argues that intention and its consequences are segments of a trajectory located in state space as it moves toward a future state. In terms of causality, he contends that intention arises necessarily influenced by past experience but is not "entirely constrained by genetic and environmental determinism." This brings us to a consideration of free will.

P8. Free Will Is an Illusion

Agency is implemented as an act of will. A current hot topic among scholars is the matter of how freely humans generate the elements of agency. The prevailing view seems to be that we delude ourselves into thinking that we make decisions freely and that all acts of agency are generated by the unconscious mind. As stated in the preceding proposition, conscious mind cannot create will or action, but only be aware of it after being generated by the unconscious mind.

A complete defense of the robot school of thought is beyond the scope of this chapter, but a comprehensive review is found in the book by Daniel Wegner (2002). Leading thinkers, such as the philosopher Patricia Churchland (2002) and the neuroscientist Michael Gazzaniga (1998) recognize the nihilistic nature of the robot conclusion but are resigned to a position of "it must be so." A recent book on this matter perpetuates the robot argument at least for many short-term intentions and asserts that the question remains open for all other intentions (Pockett et al. 2009).

The claim of illusory free will runs counter to our intuition about everyday experience. It is an extraordinary claim. In the words of Carl Sagan, "extraordinary claims require extraordinary evidence." The agency experiments conducted thus far do not provide extraordinary evidence. Moreover, a commonly used quip applies here to the criticism that the existence of free will has not been experimentally verified: "Absence of evidence is not evidence of absence." To date, definitive experiments have not been conducted. They may, in fact, be impossible to design.

A major hurdle in all discussions of "free will" is the meaning of the word "free." If one takes the stance that there is no such thing as an uncaused cause, then there can be no true freedom. However, free will is possible in the sense that one can choose among multiple options and even choose against options that provide the most positive reinforcement at the time.

Arguments over free will are commonly dominated by physical scientists. In the world view of Newtonian physics, events are governed by deterministic laws of nature, whereas the quantum mechanics world view holds that events are probabilistic. Applying such ideas to human agency seems problematic, because human

choices are made under varying states of contingency. For example, if in choosing among three viable options, each option has a certain probability of selection, the one picked is at any one instant not likely to be random, but rather its selection is determined by the existing contingencies, nor is the choice inevitably bound by a presumed probability density function. The contingencies available at some other instant are likely to be different because the context and contingencies are different, and a choice then is still not random. In short, the mind can tip the probabilities based on the context and situation at the time of choosing.

Benjamin Libet did not originate the idea that free will is illusory, but his experiments are central to modern arguments. In Libet's experiments (Libet 1985, 2006), he recorded the EEG over the motor cortex while subjects spontaneously willed to press a button. At the same time, subjects watched a clock and verbally reported the time at which they willed to press a signal button. The startling finding was that a major change in the EEG signal from motor cortex, a "readiness potential," was observed about 350 msecs before the subjects claimed that subjects said they willed the command to move. This EEG signal, discovered many years earlier by others and dubbed "readiness potential," was chosen by Libet as the index of the instant of decision. The typical interpretation of such a result is that the decision was made unconsciously and consciousness is not part of the cause. Accepting that premise, one is forced to conclude that one does not "will" such movement, but merely retrospectively confirms that there was a willed action which must have been developed subconsciously. The brain subconsciously decides to move and lets the conscious mind know what it has decided. The disturbing corollary is that one does not freely "choose" to do anything. The brain is just driven by external and internal forces to direct behavior, and one's consciousness is only around to know about it.

Belief in illusory free will drives its proponents to design experiments to disprove free will and generate specious interpretations of their data to justify the doctrine. The typical finding is that brain activity linked to a decision can be observed in one or more brain areas prior to the time the subject subjectively reports making a decision, usually to press a signal button. I have summarized the experimental design flaws and limitations elsewhere and will not repeat them here (Klemm 2010), other than to summarize the main themes:

- No one knows which neural correlates of conscious thinking are linked to decision making.
- Timing of when a free-will event occurred requires introspection, and other research shows that introspective estimates of event timing are not accurate.
- Simple finger movements may be performed without much conscious thought and certainly not representative of the conscious decisions and choices required in high-speed conversation or situations where the subconscious mind cannot know ahead of time what to do.
- Decision making is surely not a point process but rather spread out over relatively long periods as neural circuitry processes the various elements (see legend of Fig. 4.2)

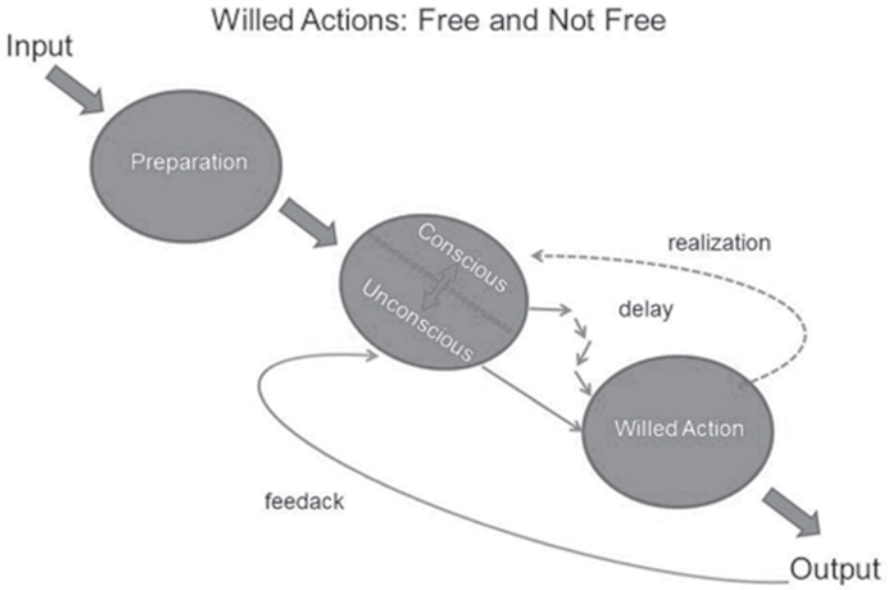

Fig. 4.2 Willed action, whether freely made or otherwise, is not a point process but rather involves serial and parallel processes. Circuits that process an intent to act must first identify a situational context, prepare relevant target populations, and launch corresponding unconscious and conscious processes involving salience and value of alternatives, recall of relevant past experience, and then selection of one of perhaps several options. The chosen action becomes an act of will which launches the appropriate output action, which often requires sustained willed engagement. The conscious mind generates choices much slower than the unconscious mind and is informed of the willed choice after some significant delay. Consequences of the willed output are fed back to the unconscious mind, and some of that feedback is available for conscious realization

- The antecedent brain activity (P2) can represent multiple processes that are not uniquely unconscious.
- Motor cortex is merely the final output path, not the origin, of the multiple elements of agency, which are processed elsewhere. As mentioned for P4, goal-directed movements are chosen, planned, and executed by networks with multiple nodes of neural processing.

At the outset of analyzing this proposition, it is crucial to identify the processes that must be occurring any time a willed action occurs, whether it is free or not; as shown in Fig. 4.2.

Willed action requires some input to the brain circuits that generate willed action. Such input may come from an external contingency or may be generate internally from some emotional drive or motivation. Next, the process involves preparation for making a decision on what to do and how to do it. This requires mental identification of the alternatives. The neural populations that represent each alternative must be triggered into competitive interaction, if the earlier winner-take-all process underlies the decision making. If movement is going to be required to implement

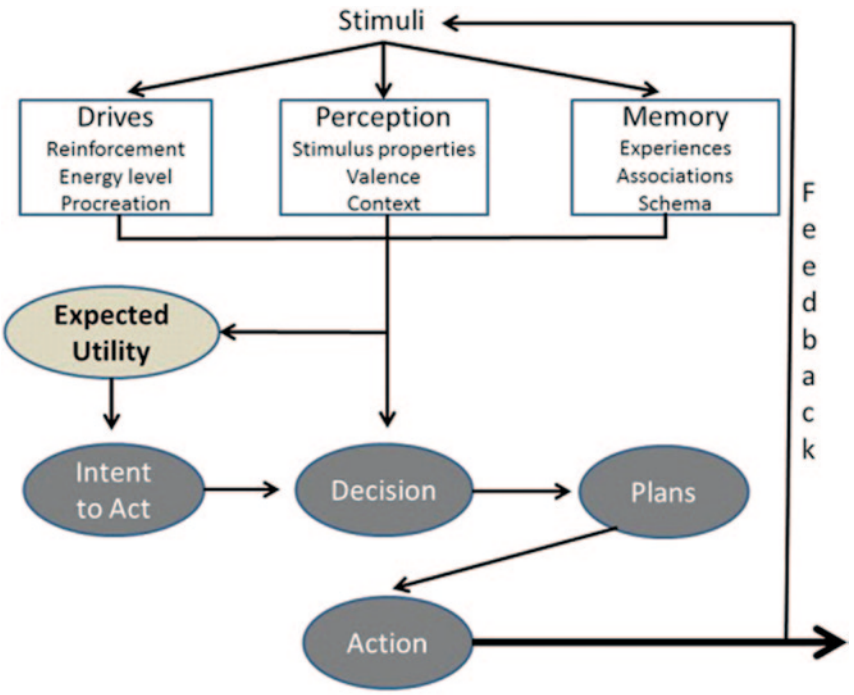

Fig. 4.3 Conscious decision-making processes. Stimuli activate drives, perceptions, and recall of memories (diagram shows some of the components of each). These processes predict whether the expected usefulness of action warrants creation of an intent to act. If intent is pursued, the decision-making process weighs the options in the context of drives, perceptions, and memories. Planning follows the decision or choice of action, which in turn generates action. Feedback monitoring during the course of action provides adjustments in the overall process

the willed action, appropriate pools of motor neurons must be engaged for readiness. All these processes will have electrical correlates long before any decision is made, and thus findings like those of Soon et al. (2008) of activated cortical areas prior to a decision should not be interpreted to indicate that a decision is made subconsciously. Some of the necessary processes in complex decision making are shown in Fig. 4.3.

How these various processes operate when a person makes a choice or decision is surely affected by the programming from past experiences. Thus, one could argue there is no free will because the agency had been programmed. Learning experiences and conscious evaluation of events program the brain for future actions by adjusting the weighting or strength of relevant synaptic connections. Such programming has the effect of setting the criteria for choices and decisions in the future. The criteria setting for future agency was accomplished in large part through conscious reflection on the processes (Fig. 4.3), which presumably provides the opportunity for a degree of freedom in imaging alternatives and evaluating their pros and cons.

This point also reinforces the evaluation of Propositions 6 and 7, which is that consciousness really does produce consequences and modulates agency.

Decisions can be made unconsciously or consciously. In the former case, those willed actions that are simple and habitual need no conscious support. This happens to be the case in most published reports that have been interpreted as proving free will to be illusory. In these studies, the willed action was typically a button press, which is not much more than a simple reflex. Certainly, pressing a button is well learned and habitual.

Among others, Rolls (2012) supports this distinction between simple and complex decisions with the conclusion about "planning ahead with multiple steps held in working memory requiring correction by higher order thoughts that may involve explicit, conscious processing." Such a view supports the idea that consciousness can produce decisions, but says little about whether or not such decisions are made freely. For that, we need some way to test the question unequivocally. Unfortunately, no one has yet devised such a test.

If we are free at all, we are free to say yes, that is to choose, or say no. Libet conceded that even his experiments supported the view that we are free to say no.

Common arguments for illusory free will derive from quantum mechanics. The problem is that there is absolutely no evidence that ordinary mental life operates on the principles of quantum mechanics. Claims to the contrary are pure speculative extrapolation of data from one domain of natural phenomena to another domain. This is a classic example of the logical fallacy known as category error.

A line of thinking that creates a bias for illusory free will is the notion that the brain operates on principles of linear causality. But almost everything the brain does is nonlinear. Freeman (2000) points out that the brain is an open system, employing self-organizing chaotic dynamics that lead to nonlinear and unpredictable behaviors. Intentionality can arise within the brain without present input because the self-organizing positive and negative re-entrant dynamics of neuronal populations enable spontaneous state transitions.

Behavior-based support for free will can be found in Murphy and Brown's (2007) book. It seems more useful to examine these issues in the practical terms of everyday living. Stimulus–response determinism does not completely describe human agency, or even agency in many higher animal species. Your dog avoids wetting the carpet because it has learned that to be unacceptable. It thus voluntarily wills to urinate elsewhere. Such observations can be taken as evidence of determinism, modified only through learning. Yet the dog can, and sometimes does, urinate in the wrong place.

Humans raise the bar of proof of determinism to a much higher level, because they can cognitively adjust focus to select stimuli to which they choose to respond. Humans have emotions that modify the qualia of experience (other higher animals do too). Humans create insight and original intentions in the absence of a stimulus.

This brings us back to earlier comments about "voluntary" movement. Behaviorists used the word to describe behaviors that appear to be more than a chain of reflexes. Willed behavior, particularly that with any degree of freedom, can certainly be more than a chain of predetermined reflexes or fixed-action patterns.

A true lack of free will would exist if a person's decisions were forced and immutable. But we all frequently make choices that we are not forced to make. True, we frequently make choices because our past experience convinces us it is wise to do so. Choices often result from conscious cost/benefit analysis. As philosopher William Barrett put it, "It is not his reason that makes man man, but rather reason is a consequence of that which really makes him man. For it is man's existence as a self-transcending self that forged and formed reason, as one of its projects" (Barrett 2000). I would add that the neurobiology of consciousness is the agency that creates self-transcendence. Are we slaves of reason because we make wise choices? If that should be true, I freely choose not to be so free.

Brains evolved to serve their own best interests. Does that inevitably strip them of the freedom of choice? No. We have all witnessed bad choices of others and perhaps ourselves when other options were equally available. If we make a choice that we know is bad for us in the long run, how does that prove we were predestined to make such a choice? We could have just perceived the odds of immediate gratification were better than the odds of negative consequences in some distant future.

Murphy and Brown put it this way in the concluding page of their book: "Neurons do not do things. You do." I agree at least with the last sentence, but not the first. You are the agent who does things, but who and what are you? Your unconscious mind and its conscious mind partner are the sum of your constructed and remembered unique set of CIPs. These come, of course, from neurons that are obviously making things happen.

Regardless of how much freedom is involved, one thing is clear: Higher-order decisions are quite complicated, involving the integration of multiple functions in multiple brain areas.

P9. Humans Create Their Own Brain Programming

Human thought and behavior are certainly programmed by genetics and epigenetic influences such as learning. But humans have the capacity to choose their environment and their learning situations. Are we to assume that such choices are always subconscious?

Whether robots or not, humans are active agents in creating their own personalities and intellectual abilities. The whole body of neuroplasticity literature testifies to the ability of experience and learning to change microanatomy and biochemistry of the brain. In that sense, humans create their own programming.

A well-established principle of neuronal development is that unused connections become lost while others become strengthened (see, for example, Changeux and Danchin 1976; Edelman 1987). Classic denervation studies by Merzenich et al. (1983a, 1983b) showed that neurons in the sensory cortex become recruited into circuits monitoring input from other nerves when their natural input nerve is removed.

These matters are reminiscent of the ongoing debate about "grandmother cells" (see Boden 2006, pp. 1205–1210). This is the idea that there are neurons dedicated

to recognition of specific objects, such as a grandmother, a specific brand of car, etc. While it is well established that there are neurons with specific dedicated functions, that does not preclude their involvement in other functions. Calculations suggest that the brain just does not have enough neurons to commit each one to a sole function. Rather, the more prevailing view is that each neuron can contribute to several or many representations and their processing.

The matter is quite different for actual movement that implements agency. While the genesis of movement from basal ganglia functions is quite obscure, there is no doubt that the final common pathway at the spinal cord level is quite specific. A given spinal motor neuron has an invariant output to specific muscles.

Humans, and to a lesser extent other mammals, are creative. A given individual creates perspectives, ideas, intentions, decisions, and actions that are new to that individual, never before witnessed or experienced. It is an open question how much of this creativity is driven subconsciously without free will. Anecdotally, many people report creative thought emerging as if out of nowhere. The recognition of such creative thought is, of course, conscious.

Few studies have examined the neurobiology of creativity. Certainly one factor must be that various neural circuits are interconnected and thus share the processing occurring within each. If these populations interact competitively, then some functions may predominate. An important recent finding is that TMS of the left, but not the right, prefrontal cortex of humans reversibly enhances creativity (Chrysikou et al. 2013). The explanation is that the prefrontal cortex is a source of executive function, which when disabled, releases creative capacity which arises from populations that were not impaired.

P10. Consciousness Is a Nerve-Impulse-Based Brain State Existing as a Being That Acts In the World

One of the original views of consciousness was an idea called the Cartesian Theatre. It is named after Rene Descartes (1596–1650), a French philosopher, who argued that there is a real distinction between physical brain and nonphysical mind. He considered that brain activity is presented as episodes on the stage of a "Cartesian Theatre" for viewing by a virtual little man. Though this idea is ridiculed by most scholars, it resurfaces in the writings of prominent neuroscientists who hold that consciousness is an "observer," not a maker of events (see Gazziniga 1998).

Daniel Dennett has taken a lead in demolishing the Cartesian Theatre metaphor by suggesting a "multiple drafts" metaphor for consciousness (Dennett 1991). This view regards consciousness as a succession of multiple streams of conscious (drafts), which in essence are snapshots of different portions of the stream of consciousness. However, this view does not explain what creates those states are or what causes them.

All modern theories assume the requirement for a lot of new cortical circuitry to support the added dimension of consciousness to the SoS. One basic idea,

sometimes referred to as "Global Neuronal Workspace," holds that many of cortical neurons across all parts of the cortex are widely interconnected and collectively constituted a global workspace (reviewed by Gray 2004). Gray, however, points out that just labeling consciousness this way does not make the "hard problem" of explaining consciousness go away.

There is certainly plenty of anatomical evidence that our cortexes are built in ways that could readily create a global workspace (see Douglas and Martin 2004). The workspace idea holds that localized networks of neuronal processes, as might be going on in the emotional part of a human brain, or in the auditory part, visual part, etc. compete for attention. That is, these multiple parallel operations compete for access to the workspace.

This is a "bottom up" idea, where consciousness (at the top) is a recipient of input that has gained a competitive advantage for conscious "viewing." But why could conscious SoS not be a "top down" process, where the conscious mind "looks in" on local processes and selects which ones it was to pay attention to and modify?

What seems missing in the global workspace view is that it does not seem personalized. Everybody's global workspace is unique. Each has been sculpted by unique genetics and personal experience.

We know that consciousness comes from and resides in the brain. It is either a state of function or a "being" with functions of self-awareness, awareness of non-self, and perhaps the ability to generate intentions, choices, and actions. If it is a "being," consciousness may likely be constituted from a unique global set of nerve impulse patterns flowing in an extensive neural network. When consciousness temporarily ceases, as in sleep, those patterns must be stored in the respective synapses that can generate the same patterns upon reawakening. As such, we could think of consciousness as an avatar, generated by the brain as a set of CIPs to act on behalf of the best interests of the brain and body (Klemm 2011a, 2011b).

Irrespective of which way is up or down, consider the possibility that the conscious self is represented by CIPs, as is the mode in which brains represent everything else. Now consider the possibility that the CIP representation of the conscious self is the equivalent of an avatar that acts in the world on behalf of the brain and body. This consciousness CIP incorporates a portion of the global workspace of the whole brain and is therefore integrated with what goes on in unconscious processing. The SoS CIPs thus constitute a being.

People who play computer games may know about avatars. These are computer-generated proxies for the gamer. The avatar does things in the game in response to commands from the player. A good example is the increasingly popular Web environment known as Second Life, in which players create their own avatars and live vicariously through the avatar in the virtual world.

Unlike computer avatars, however, the brain's avatar might do things on its own initiative to serve the brain and body's best interests. Perhaps, the avatar has an interest of its own, a mind of its own so to speak. Moreover, the biological avatar gets to decide what it is that is in the best interest and supervise the actions to accomplish it.

This avatar notion is reminiscent of the old third-century idea, typically deemed foolish, that the brain has a little person inside that is a scale model of the body. In modern terms, however, this homunculus, as it came to be called, is used as a way to think about how the body is mapped in sensory and motor cortex. But the homunculus is more than that. It is a CIP representation of the body and what goes on inside and outside of the body, all referenced to the SoS, which itself is a CIP representation of the "little person." The brain could create a conscious homunculus in the form of an avatar that it deploys to act on behalf of the embodied brain in ways not otherwise possible.

I recently found a related perspective in the literature on phantom limbs (Melzack 1989). Melzack posited a body-self "neuromatrix," which he defines as a spatially distributed network of neurons that acts a whole to generate a nerve-impulse-based neural signature pattern of the self. The avatar idea extends neuromatrix beyond its emphasis on knowledge about the body to regard the conscious SoS as a being with agency.

Avatars sense, evaluate, decide, and initiate and direct action. What is sensed by the avatar? First and foremost, it senses its own identity and constructs it as a being in the form of CIPs. It is the sense of "I." The CIP representations of stimuli are registered and integrated into the CIP representations of the conscious avatar. The avatar also senses much of what the brain is thinking, such as beliefs, wishes, decisions, plans, and the like. Moreover, the events in consciousness can teach the unconscious brain, whether it is in terms of specific sensations, cognitive capabilities, motor skills, ideas, attitudes, or emotions.

"I Avatar" thinks of itself in some of the biological ways that subconscious and non-conscious minds think of themselves as existing as part of the embodied brain. But the avatar thinks consciously of itself, though not independently from unconscious minds. If conscious being is constituted as a self-aware avatar, then of course it thinks of itself consciously. The point is that thinking of oneself as an avatar and being an avatar are equivalent. Recall the famous saying of philosopher Rene Descartes,

> I think, therefore I am.

Any time we are awake, the avatar being is active—deployed online so to speak. Think of this as a computer analogy: When operating in RAM, the avatar is online and available to exert its functions. When the avatar is shut down, as in going to sleep for example, the avatar goes off-line and saves its CIP files on hard disk. In biological systems, the hard disk is stored in the neuron terminals and synapses of the circuits that hold the memory of the self and the capacity for rebooting the self. The self may have undergone some subtle changes with the day's experiences, and updating the modified self in long-term memory is one of the functions of sleep.

The avatar exists "online" as a unique, individual-specific, set of global CIPs that constitute a "being." When "off-line," the avatar exists in a corresponding memory of self, stored in the synapses of the circuits that generate the avatar when online.

4.5 Needs for Future Research

Most neurobiological research has not been conducted in the context of understanding agency as such. The needed future research in this area should aim first at identifying neural correlates of the various elements of agency (recall Fig. 4.3), and where possible aim at distinguishing unconscious and conscious agency. What little past research exists has often been based on the simplistic assumptions that do not adequately account for the various elements of agency and their seamless integration with each other. In terms of the neurobiological approaches that might identify specific correlates of these agency elements include the following.

4.6 Top-Down/Bottom-Up Issues

The intuitive explanation for agency would seem to invoke top-down, executive control. Stimulus-driven responses reflect bottom-up neural processes, whereas a person's choices may reflect top-down processes. Unfortunately, distinguishing the two is notoriously difficult (Folk et al. 1992; Ogawa and Komatsu 2006). A likely reason is that self-organizing systems, as mentation certainly can be, need not rely on hierarchical functions.

Top-down control obviously accounts for agency in invertebrates with command-neuron systems. But in higher animals, particularly humans, there is no "top" in the sense that brain circuitry is highly interconnected and activity at any level can be communicated in direction, whether it be up or down or lateral. Moreover, such interacting networks readily oscillate, and their interactions may self-organize. The areas of the brain most conspicuously prone to oscillation are the visual cortex (alpha rhythm), thalamus cortex (spindles), the hippocampus (theta activity), and the neocortex in general (gamma waves). These brain areas are richly interconnected and their differing oscillatory frequencies must exhibit shifting degrees of synchrony or phase locking. Such a mechanism of cooperative processing calls into question any notions of top-down or bottom-up processing.

All cortical circuits can generate oscillating self-organized activity, no doubt as a result of the intrinsic circuit design of cortical columns. Thus, agency need not arise from external output, but can emerge from the ongoing emergence of memory representations in linked circuitry. The degree of coherent coupling among oscillatory activity provides a way for self-organized intents and decisions to emerge.

That being the case, it is important to reexamine our biases about executive control. For example, we generally think of the motor cortex as "ordering" action of specific muscles. This underlies some of the flaws in interpreting free-will experiments based on the experimental designs of Libet and his followers, as discussed earlier. The reality is that the motor cortex does not operate in isolation. It is coupled with many oscillating systems, both cortical and subcortical. Whether and when motor cortex cells discharge "commands" depends on prior happenings in the

multiple circuits with which it is coupled. Intentions and decisions that occur before the motor cortex delivers an output are being processed elsewhere, long before action is commanded.

4.7 New Strategies

1. Focus on causal neural correlates of the elements of agency. I suggest that a starting point for new research is the need to identify more completely the neural correlates of agency, distinguishing the correlate differences between conscious and unconscious states. Moreover, little research has aimed at identifying neural correlates with specific elements of agency. But such research is likely to prove fruitful in light of the few studies that have been thus far reported.
2. Use strategies that reduce confounding variables. I have mentioned some of the confounding variables and sources of error in the traditional experiments on free will. One example of the kinds of experimental design needed in future research is to use stimuli that are constant, independently of any cognitive function. My laboratory's study of ambiguous figures had the advantage that the stimulus for a given bi-stable image was constant—the only thing that could change was the conscious percept. We found marked coherence shifts in all human EEG frequency bands during the second when subjects reported an "aha" moment when they perceived the alternate image (Klemm et al. 2000). For any of the ten ambiguous figures used, a subject had a default percept, but through attending to certain features of the image and force of will (focused attention on specific portions of the image) they would perceive the alternate image (as, for example, first seeing a vase and then realizing that it could also represent two faces in profile). The coherence shift occurred with all of the images, so the coherence represented the process of willed percept and decision, not the specific image.

Import aspects of ambiguous figures that we never got to examine could have great bearing on the process of intention and decision making. First, there is need to compare the neural correlates for each percept as well as those during the switching process. You could also compare correlates of agency, such as when a subject willfully resists switching percepts versus states where the subject willfully switches intentionally back and forth between percepts. All such states are clearly representative of consciousness; they cannot occur during unconsciousness.

4.8 Tactics

1. Observe fMRI correlates with different elements of agency. They should differ.

fMRI is a powerful and increasingly used tool in neuroscience research. While little of this approach has been used in the context of studying agency, this chapter has cited some relevant recent fMRI studies. More are likely to be forthcoming.

Brain scans can identify regional differences in brain activity during different kinds of cognitive processes. This has already served the purpose of showing that some elements of agency occur in distinctively different brain areas. Scans can also help identify the sequencing of different elements of agency.

On the other hand, fMRI does have serious limitations. It measures metabolism, not the brain's information carriers (CIPs or field potentials). It is not always clear what an increase in metabolism means in terms of information processing. Moreover, regional decreases in fMRI are seldom studied, yet they certainly reflect some kind of significant change in neural processing.

2. Observe how elements of agency are affected by reversible blockage of specific cortical sites with TMS.

The beauty of TMS is that it can reversibly shut down activity in underlying cortical areas. Such blockade of defined cortical areas might elucidate their role in specific elements of agency. Some promising research of this kind was cited herein and more can be expected.

This technology also facilitates the exploitation of the strategy of evaluating staged recovery processes after key brain areas have been inactivated. This strategy was used to good effect in the research of Teitelbaum, who in his time had only the technical ability to create permanent lesions (Marshall and Teitelbaum 1977).

3. Evaluate oscillatory field potential coherence changes associated with different elements of agency.

Gamma field potential oscillations are highly involved in active thinking, as has been demonstrated in numerous species and laboratories. It is likely that gamma oscillations, especially their coherence between and among brain areas, participate in intention genesis and decision making.

Coupling of networks oscillating at different frequencies can also be achieved. Information contained within one oscillating network could be communicated to another oscillating network if the target network has a faster oscillation phase locked to the input and the summed activity of the two oscillators sufficiently depolarizes the neurons (Buzsáki 2006). Unfortunately, very few researchers test for coherence of different networks, but rather limit their investigations to coherence of a single frequency band in multiple brain regions. I know of no studies that have examined phase coupling of different frequencies, either within or between brain areas, under conditions of agency.

Both the hippocampus and entorhinal cortex display theta oscillations when the brain is activated and likely to be engaged with and exploring the environment. In both brain areas, the voltage power of gamma activity increases in the presence of theta oscillations and varies dynamically with the relative phase relationships (Fig. 4.4; reviewed by Buzsáki 2006).

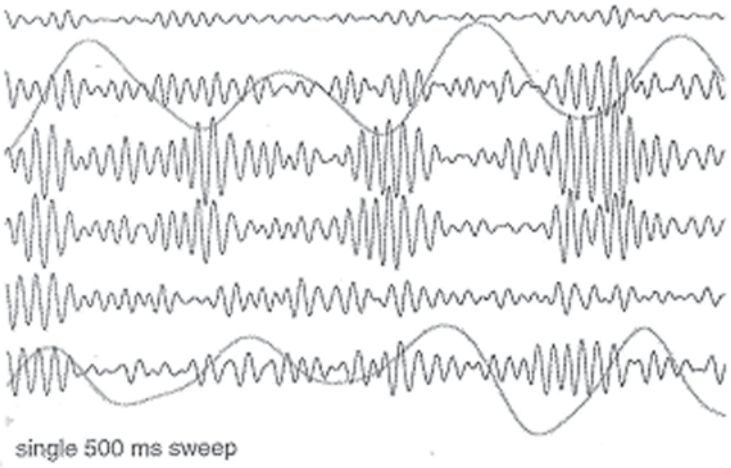

Fig. 4.4 Phase coupling between entorhinal theta field potentials and higher frequency of simultaneously occurring gamma waves in the rat. Note that gamma amplitude increases in the downward theta phase. Frequencies were separated by electronic filtering. (Reprinted with permission from Buzsáki 2006)

Such frequency modulation alters the throughput of nerve impulses during the theta troughs and may package impulse "messages" in ways that preserve the message as well as make it more "readable" by targets.

Similar modulation of neocortical alpha waves occurs in the human gamma power (Freeman et al. 2003). We can assume that the nerve impulses represent cooperativity in the circuits in which they propagate, packaged in clusters of a fraction of a second and dissolving in the down cycle to make way for the next cluster. Such a process would likely support agency, as it progresses sequentially from intention genesis, to decision, to planning, and to execution.

The same mechanisms could operate with regard to incorporating memory of prior learning into real-time agency. Memory, along with salience and reward valence, is central to agency. The stored synaptic strengths of networks storing the memory, once activated by recall, can provide a similar degree of stochastic resonance as occurs with novel contingencies.

Phase locking of multiple neural oscillators is typically interpreted in the context of perception and in particular the "binding problem" (reviewed by Metzinger 1995). Since most elements of agency also engage diverse distributed populations, should not the idea of binding apply also to the genesis of agency? This remains an undeveloped area of research.

4. Evaluate state-space field potential correlations with the elements of agency.

In nonlinear complex dynamic systems such as the brain, discriminating causation from correlation is particularly difficult. Scientists know that correlation is neither necessary nor sufficient to establish causation, yet they frequently violate this basic

tenet of logic. Also, they may conclude that lack of causation is not established by lack of correlation.

Computational modeling of causation is not usually done in the context of human agency, but perhaps it should be. Relevant computational techniques are emerging. For example, Sugihara et al. (2012) report an approach based on nonlinear state-space reconstruction that can detect when correlation is a cause—basically, a process that identifies whether two variables share a common dynamical system. Their basic calculations measure the extent to which the historical record of *Y* time series values can reliably estimate states of *X*. This happens only if *X* causes *Y*. Cross-mapping computations check for a reliable relationship between the sets of data points in the attractor state-space surfaces of *X* and *Y* variables. If such relationships exist, *Y* can estimate *X* and vice versa. This method was applied to ecological data, but in principle the approach could apply to dynamic changes in electrographic signatures of human agency.

5. Record impulse activity and perform combinatorial analysis of the temporal patterns in defined circuitry such as cortical columns

The SoS and its agency is represented in the brain by patterns of nerve impulses propagating in circuitry (CIPs) that could contain a combinatorial code of impulse activity, a possibility that is ripe for investigation. The concept is that circuits of interacting neurons can have specific properties of their own that are not evident in any one-member neuron. Moreover, the phase and frequency relationships of the CIPs are likely representations of sensation of both ordinary senses and the capacity for self-initiated agency. These representations no doubt change dramatically as the capacity for consciousness develops. The circuit elements of cortical columns are known, and new microelectrode technology and combinatorial mathematics make this area of research an area whose time has come.

4.9 Why Agency Research Matters

At least some of the axioms and propositions should spark interest and controversy in the research community. Indeed, this is already true in several cases, such as in research on whether consciousness has causal effects, whether free will exists, how decisions are made, whether agency is top-down or bottom-up, or whether these are even useful conceptual frameworks.

These issues of axioms and propositions will perhaps be seen by many as academic and arcane, but agency is a very practical matter in the real world of human behavior. Neurobiological research has not yet provided satisfying answers to a range of questions about why people act the way they do:

- Why are some people more assertive than others?…some more active, others more passive?
- Why are there individual differences in initiative, and in some people, extending to impulsivity?

- Why are some people more courageous than others?
- Why are some people risk takers and others risk averse?
- Why are aggressive predilections expressed passively in some people but actively in others?
- Are there genetic or neural correlate markers for the range of active to passive personalities?
- What environmental factors influence the ontogenetic development along the agency continuum?
- Can a given agency personality type be modified by education, training, or behavioral therapy?
- What is the relationship of mental health to agency?
- Does aging affect agency, and, if so, why and how?

Perhaps research addressing these axioms and propositions will lead to new propositions, but more importantly lead us to find better ways for people to control their own behavior in more appropriate and effective ways.

4.10 Conclusions

Animals at all levels of complexity must act appropriately in order to survive. At the simplest level, such actions are intrinsic, automated, and unconscious. At slightly higher levels, such actions are responses to stimuli. At still higher levels, actions arise from such agency elements as inherent drives, perception, memory recall, expected utility, intent to act, decision, plans, and finally action. These action elements may be unconscious, conscious, or some mixture of both.

The neural basis for agency exists in the circuitry connections of neurons which are most fundamentally hardwired in simple brains with relatively few neurons. But in higher animals, the superabundance of networks within networks allows for dynamic adjustment of circuitry to enable adaptive adjustment of the elements of agency. In all circuits, the information carrier is the spatiotemporal patterns of nerve impulses.

Therefore, if we know the circuitry and their impulse patterns, we may know the cause of any element of agency. We also need to know how the CIPs of higher animals seamlessly mediate the sequence of elements of agency to produce a final action.

Two strategies are proposed: (1) focus on causal neural correlates of specific elements of agency and (2) employ experimental designs that reduce confounding variables. Among the specific tactics suggested are: (1) observe how fMRI correlates with different elements of agency, (2) observe how elements of agency are affected by reversible blockage and recovery of specific cortical sites with TMS, (3) evaluate oscillatory field potential coherence changes associated with different elements of agency, (4) evaluate state-space field potential correlations with the elements of agency, and (5) record impulse activity and perform combinatorial analysis of the temporal patterns in defined circuitry such as cortical columns.

References

Banich, M. T., Mackiewicz, K. L., Depue, B. E., Whitmer, A. J., Miller, G. A., & Heller, W. (2009). Cognitive control mechanisms, emotions and memory: a neural perspective with implications for psychopathology. *Neuroscience and Biobehavioral Reviews*, *33*, 613–630.

Barik, J., Marti, F., Morel, C. et al. (2013). Chronic stress triggers social aversion via glucocorticoid receptor in dopaminoceptive neurons. *Science, 339,* 332–335.

Barrett, W. (2000). *Irrational man. A study in existential philosophy* (p. 279). New York: Anchor Books.

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry, 165,* 969–977.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., & Wiener, S. I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron, 66,* 921–936.

Boden, M. A. (2006). *Mind as machine. A history of cognitive science* (Vol. 2., pp. 1205–1210). Oxford: Oxford University Press.

Briggman, K. L. et al. (2006). Optical imaging of neuronal populations during decision-making. *Science, 307,* 896–901.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science, 340,* 95–98.

Buzsáki, G. (2006). *Rhythms in the brain*. New York: Oxford University Press.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304,* 1926–1929.

Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Review of Neuroscience, 16,* 223–243.

Changeux, J.-P., Danchin A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. Nature 264, 705–712.

Chrysikou, E. G., Hamilton, R. H., Coslett, H. B., Datta, A., Bikson, M., & Thompson-Schill, S. L. (2013). Noninvasive transcranial direct current stimulation over the left prefrontal cortex facilitates cognitive flexibility in tool use. *Cognitive Neuroscience*. doi:10.1080/17588928.2013.768221.

Churchland, Patricia S. (2002). Self-representation in nervous systems. *Science, 296,* 308–310.

Dennett, D. (1991). *Consciousness explained*. Boston: Little Brown.

Di Pisapia, N. (2013). Unconscious information processing in executive control. *Frontiers in Human Neuroscience*, Jan. 31. doi:10.3389/fnhum.2013.00021.

Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience, 27,* 419–451.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic.

Evarts, E. V., & Tanji, J. (1976). Reflex and intended responses in motor cortex pyramidal tract neurons of monkey. *Journal of Neurophysiology, 39,* 1069–1080.

Filimon, F., Philiastides, M. G., Nelson, J. D., Kloosterman, N. A., & Heekeren, H. R. (2013). How embodied is perceptual decision making? Evidence for separate processing of perceptual and motor decisions. *The Journal of Neuroscience, 33*(5), 2121–2136. doi:10.1523/JNEUROSCI.2334-12.2013.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044.

Freeman, W. J. (2000). *How brains make up their minds*. New York: Columbia University Press.

Freeman, W. J., Burke, B., & Holmes, M. (2003). Aperiodic phase resetting in scalp EEG of beta-gamma oscillations by state transition at alpha-theta-rates. *Human Brain Mapping, 19,* 248–272.

Gallivan, J. P., McLean, D. A., Flanagan, J. R., & Culham, J. C. (2013). Where one hand meets the other: Limb-specific and action-dependent movement plans decoded from preparatory signals

in single human frontoparietal brain areas. *The Journal of Neuroscience, 33*(5), 1991–2008. doi:10.1523/JNEUROSCI.0541-12.

Gazzaniga, M. S. (1998). *The mind's past* (201 p.). Berkeley: University of California Press.

Gray, J., (2004). Consciousness: Creeping up on the Hard Problem. Oxford: Oxford University Press.

Guitart-Masip, M., Barnes, G. R., Horner, A., Bauer, M., Dolan, R. J., & Duzel, E. (2013). Synchronication of medial temporal lobe and prefrontal rhythms in human decision making. *The Journal of Neuroscience, 33*(2), 442–451.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5,* 382–385.

Ham, T., Leff, A., de Boissezon, X., Joffe, A., & Sharp, D. J. (2013). Cognitive control and the salience network: An investigation of error processing and effective connectivity. *The Journal of Neuroscience, 33*(16), 7091–7098.

Klemm, W. R. (1976). Physiological and behavioral significance of hippocampal rhythmic, slow activity ("theta rhythm"). *Progress in Neurobiology, 6,* 23–47.

Klemm, W. R. (1990). The behavioral readiness response. In W. R. Klemm & R. P. Vertex (Eds.), *Brainstem mechanisms of behavior* (pp. 105–145). New York: Wiley.

Klemm, W. R., Li, T. H., & Hernandez, J. L. (2000). Coherent EEG indicators of cognitive binding during ambiguous figure tasks. *Consciousness and Cognition, 9,* 66–85.

Klemm, W. R. (2001). Behavioral arrest: In search of the neural control system. *Progress in Neurobiology, 654,* 453–471.

Klemm, W. R. (2010). Free will debates: Simple experiments are not so simple. *Advances in Cognitive Psychology, 6*(6), 47–65.

Klemm, W. R. (2011a). *Atoms of mind. The "ghost in the machine" materializes*. New York: Springer International.

Klemm, W. R. (2011b). Neural representations of the sense of self. *Advances in Cognitive Psychology*, *7*(7), 16–30.

Klemm, W. R. (2011c). Why does REM sleep occur? A wake-up hypothesis. *Frontiers in Neuroscience, 5*(73), 1–12. doi:10.3389/fnsys.2011.00073.

Klemm, W. R. (2012). Sense of self and consciousness: Nature, origins, mechanisms, and implications. In A. E. Cavanna & A. Nani (Eds.), *Consciousness: States, mechanisms and disorders*. Hauppauge: Nova Science. https://www.novapublishers.com/catalog/product_info.php?products_id=38801.

Klemm, W. R. (2013). Just do it. *Psychology Today*. http://www.psychologytoday.com/blog/memory-medic/201303/just-do-it–0.

Klemm, W. R. (2014). The brain's odyssey. New York: Prometheus. (In press)

Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimullus oscillatory activity over motor cortex reflects perceptual expectations. *Journal of Neuroscience, 33*(4), 1400–1410. doi:10.1523/JNEUROSCI.1094-12.2013.

Libet, B., & Commentators (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529–566.

Libet, B. (2006). The timing of brain events: Reply to the "special section" in this journal of September 2004. In Susan Pockett (Ed.), *Consciousness and Cognition, 15*(3), 540–547.

Luk, C.-H., & Wallis, J. D. (2013). Choice coding in frontal cortex during stimulus-guided or action-guided decision-making. *Journal of Neuroscience, 33*(5), 1864–1871.

Mandler, George (2003). Consciousness: Respectable, useful, and probably necessary. In B. J. Baars et al. (Eds.), *Essential sources in the scientific study of consciousness* (pp. 15–33). Cambridge: MIT.

Marshall, J. F., & Teitelbaum, P. (1977). New considerations in the neuropsychology of motivated behaviors. In *Handbook of psychopharmacology* (pp. 201–229). New York: Springer.

McMains, S., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience, 31*(2), 587–597.

Melzack, R., (1989). Phantom limbs, the self and the brain. Canad. G. Psychol. 30, 1–16.

Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J., & Felleman, D. J. (1983a). Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys. *Neuroscience, 10,* 639–665.

Merzenich, M. M., Kaas, J. H., Wall, J. T., Nelson, R. J., Sur, M., & Felleman, D. J. (1983b). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience, 8,* 33–55.

Metzinger, T. (1995). Faster than thought. Holism, homogeneity and temporal coding. In T. Metzinger (Ed.), *Conscious experience* (pp. 425–461). Thorverten: Imprint Academic.

Murphy, N., & Brown, W. S. (2007). *Did my neurons make me do it?* New York: Oxford University Press.

Niwa, M., Jaaro-Peled, H., Tankou, S. et al. (2013). Adolescent stress-induced epigenetic control of dopaminergic neurons via glucocorticoids. *Science, 339,* 335–339.

Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research, 173,* 1–13.

Paz-Alonso, P. M., Bunge, S. A., Anderson, M. C., & Ghetti, S. (2013). Strength of coupling within a mnemonic control network differentiates those who can and cannot suppress memory retrieval. *Journal of Neuroscience, 33*(11), 5017–5026.

Pockett, S., Banks, W. P., & Gallagher, S. (Eds.) (2009). *Does consciousness cause behavior?* Cambridge: MIT.

Povinelli, D. J., & Giambrone, S. (2001). Reasoning about beliefs: A human specialization? *Child Development, 72,* 691–695.

Rivard, G., & Klemm, W. R. (1990). Sample contact required for complete bull response to oestrous pheromone in cattle. In D. W. MacDonald, D. Muller-Schwarze & S. E. Natynczuk (Eds.), *Chemical signals in vertebrates V* (pp. 627–633). Oxford: Oxford University Press.

Rolls, R. T. (2012) Willed action, free will, and the stochastic neurodynamics of decision-making. *Frontiers of Integrative Neuroscience*. doi:10.3389/fnint.2012.00068.

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science, 340,* 216–219.

Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., & Dalgleish, T. (2013). Training the emotional brain: improving affective control through emotional working memory training. *Journal of Neuroscience, 33*(12), 5301–5311.

Soon, Chun Siong et al. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 13 April. doi:10.1038/nn.2112.

Soutschek, A., Taylor, P. C. J., Müller, H. J., & Schubert, T. (2013). Dissociable networks control conflict during perception and response selection: a transcranial magnetic stimulation study. *Journal of Neuroscience, 3*(13), 5647–5654.

Studer, B., Apergis-Schoute, A. M., Robbins, T. W., & Clark, L. (2012). What are the odds? The neural correlates of active choice during gambling. *Frontiers in Neuroscience*. doi:10.3389/fnins.2012.00046.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S. (2012). Detecting causality in complex ecosystems. Science. 338, 496–500.

Taylor, J. G., & Alavi, F. N. (2003). A global competitive network for attention. In J. Baars et al. (Eds.), *Essential sources in the scientific study of consciousness* (pp. 1035–1058). Cambridge: MIT.

Verhagen, L., Dijkerman H. C., Medendorp, W. P., Toni, I. (2013) Hierarchical organization of parieto-frontal circuits during goal-directed action. J. Neurosci, 33(15), 6492–6503.

Von Stein, A., Chiang, C., & König, P. (2000). Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Sciences, 97*(26), 14748–14753.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge: MIT.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. Hassin et al. (Ed.), *The new unconscious*. Oxford: Oxford University Press.

Chapter 5
Agency in Life

Jing Zhu

In "Neurobiological Perspectives of Agency" (Chap. 4 this volume), neuroscientist W. R. Klemm provides a comprehensive survey of the neural basis of agency, summarized in ten axioms and ten propositions. According to Klemm, the former, by and large, are consensuses in current neuroscience, whereas the latter are still controversial and open to debate. Understanding agency matters, for "agency is a very practical matter in the real world of human behavior" (Chap. 4, p.?). Thus, it is not surprising that agency has also been a subject of study in many other disciplines aside from neuroscience, including philosophy, psychology, sociology, and computer science (artificial intelligence and robotics).

Neuroscience is dedicated to exploring the neural correlates of elements of agency. A prior issue that needs to be addressed is how the concept of agency is to be understood properly, which in turn shapes how the elements of agency can be specified, and where and how to look for their neural correlates. On the other hand, data and findings from neuroscience may well inform our general understanding of agency.

In this chapter, I shall focus on the concept of agency, especially the structures of agency. Drawing on some insights offered in Klemm's chapter, I attempt a broader conception of agency than which is usually conceived, putting agency in a creature's ecological niche and whole life realm, to highlight the significance of consciousness in human agency.

5.1 Structures of Agency

As Klemm remarks (Chap. 4, p.?):

> Understanding agency begins with the recognition that agency typically consists of multiple elements, such as the generation of intent, identification of contextual salience, recall of relevant memories, evaluation of choice options and their anticipated consequences,

J. Zhu (✉)
Sun Yat-sen University, Guangzhou, China
e-mail: zhujing6@mail.sysu.edu.cn

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_5

> decision making, planning, implementing the action, and real-time monitoring and adjustment of the implementation.

Agency arises from the interactive processes between a creature and its environment. The simplest form of agency may be embodied in reflex action, "where sensory input more or less induces an automated response, often with only a minimal amount of processing" (Chap. 4, p.?). Higher animals with true brains can create purposes and make decisions that fulfill purpose. Human brains can create conscious intentions and plans, and have them carried out.

Klemm suggests that human agency consists of the following elements:

> intend, remember, value, decide, prepare/plan, and act.

Each of these elements of agency is distinctly different. The task of neuroscience is to identify the neural correlates of the elements of agency. I think this strategy is correct and promising. However, what I want to emphasize here is that not only do we need to specify the elements of agency but also we need to understand the structures of agency, namely, how the elements of agency are related to each other and organized. For example, what are the relations between deciding, intending, and planning? How do distal or future-directed intentions interplay with proximal or present-directed intentions? How do values affect an agent's decision making and intentions? How do decisions, intentions, and plans work to initiate, guide, and control actions? These issues are important in that they are essential to our understanding of the work of agency.

The philosopher Michael Bratman (1987, 1999, 2007) developed an account of human agency centered on the notion of planning. According to Bratman's planning theory of agency, intentions are characteristically elements of larger, partial plans of action, and these plans play basic coordinating, organizing roles at a time and over time. Associated with these roles are distinctive rational pressures on intentions for consistency and coherence at a time, and stability over time. Intentions typically involve a sort of psychological or practical commitment: An agent who holds an intention is settled upon or committed to engaging in the course of the intended action, either in the future or at present, although the commitment is not irrevocable. Intentions tend to resist reconsideration and revision. Once an intention is formed, sometimes as a result of deliberative decision making, slight changes of circumstance will generally not lead the agent to reconsider the intention or reopen the deliberative process. Intention thus "has a characteristic stability or inertia" in human practical reasoning and action (Bratman 1987, p. 6). These two distinctive features of intention, namely, the settledness and inertia of intention, help to explain how intentions play their characteristic roles in shaping people's practical reasoning and in supporting interpersonal and intrapersonal coordination. The expectation of the behavior of other agents based on their intentions is central to social coordination. When we are rationally planning for our future actions, we need to take our prior intentions seriously.

For example, last week I saw a notice calling for participants by a psychology research laboratory. When I decided to participate in an experiment in cognitive

neuroscience and made a phone call for an appointment, I formed an intention to take part in the experiment as a subject, at an agreed time this week. I formed the intention because this is consistent with the values I hold, say, I value their research effort and hope to make my contribution, however slight. And I also wanted to know, out of my curiosity, how an experiment in cognitive neuroscience is typically carried out. This intention is a part of my plan for this week. It is consistent with, or at least, not in conflict with my other prior intentions or parts of my plan. If it turns out that, for some reasons, I cannot fulfill the appointment, I may call to change for another time or even to cancel it. But if no such thing happens, I shall stick to the appointment accordingly. Thus, this intention, externalized through a phone call appointment, serves as a commitment between me and the laboratory people, which coordinates our plans and actions thereafter. It also works as a constraint on my schedule: I shall not form any (further) plans or intentions in conflict with it.

From this example, we can see that planning is indeed a crucial part of human agency. We have various plans. Some are for the far future, such as those about career, marriage, and lifestyle, whereas some are for the near future, such as those about important things to be done tomorrow or next week. Some may be quite specific in content, while some are vague. Whereas some are complete, most plans are partial, with many details to be figured out in due course. Planning plays a key role in organizing values, decisions, intentions, and actions, and designates, in Bratman's words, the temporal extension of human agency.

In comparison with the elements of agency that Klemm has identified, structures of agency may seem more elusive, especially in looking for their neural correlates in experimental settings. But they shall not be neglected. Interpretations of experiment data and laboratory findings need be cautious to project to real-life cases, where the elements of agency make sense more likely in light of structures of agency.

5.2 Consciousness and Agency

This morning after waking up in bed, I recalled that I have an appointment this afternoon with the psychology laboratory to participate in a cognitive neuroscience experiment. This appointment was made last week. As there was nothing more urgent forcing me to break the commitment, I decided to go to the laboratory. After giving me the instructions on how to perform in the experiment and ensuring my understanding of said instructions, the experimenter put an electroencephalogram (EEG) equipment on my head to record electrical activities along the scalp, and an electromyography (EMG) detector on my right wrist to record its movement. I was instructed to perform quick flexion of my right wrist on my own initiative, at any time I felt the "urge" or desire to do so. For each trial, I was required to watch the "clock position" of a spot of light revolving in a circle on the face of a screen, which run much faster than a normal clock, then, after moving my wrist, recall the clock time of the first awareness of the urge to move. About several 100 trials had been made in the experiment, before I was paid a small amount of money and left the laboratory.

The above scenario is a somewhat simplified description of the well-known Libet experiment (Libet et al. 1983). In Libet's experiment, it was found that the average time for the subject's awareness of the urge to move was 200 ms before the activation of the muscle. But a slow, negative potential shift recorded by the EEG on the scalp, called Bereitschaftspotential or readiness potential (RP), was found to have an onset for its main negative rise at about 550 ms before the actual motor movement began. Thus, a subject was typically aware of the urge or intent to act 350 ms later than the RP had already emerged in performing a self-initiated flexion.

An orthodox interpretation of Libet's experiment is that our unconscious brain processes and brings about voluntary actions. Our voluntary acts are "initiated by unconscious cerebral processes before conscious intention appears" (Libet 1985, p. 529). Onsets of RPs, which are unaware to humans, regularly begin at least several 100 ms before conscious intention appears. Hence, spontaneous voluntary movements are actually initiated unconsciously by cerebral processes, instead of conscious intentions or volitions as conventionally conceived. "The brain 'decides' to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place," which entails that "some neuronal activity associated with the eventual performance of the act has started well before any (recallable) conscious initiation or intervention is possible" (Libet 1985, p. 536).

Along with Klemm, I go against the standard interpretation of Libet's experiment (see Zhu 2003 for an earlier attempt for an alternative). Here are some observations that should be noted, but have been widely neglected in discussions of the Libet-style experiments:

First, when I entered the psychology laboratory, I was awake and conscious, fully aware of what I was doing. I formed an intention to participate in the experiment last week, which had been a part of my overall plan for this week. This intention had played a role in coordinating, guiding, and controlling my activities over time since its formation, till the time of its execution, unless I had changed my mind due to other causes or had totally forgotten it. Had I not had this intention, or had I not been awake and conscious, and had not wanted to execute this intention or plan, I would not have walked myself into the laboratory. I found myself standing in the laboratory with no surprise. If all my seemingly voluntary behaviors are initiated by my unconscious cerebral processes, whose precedent intentions or urges I am aware of only after these behaviors have already been in preparation or execution processing, it would be a real miracle that these behaviors are deployed largely in accordance with my intention formed one week ago.

Second, when I decided to participate in the experiment last week, I was settled upon cooperating with the experimenters to try to complete the task. This commitment became entrenched as I entered the laboratory: Had I not wanted to be cooperative, I would most likely have not presented myself to the laboratory and would better have had something else to do. Therefore, I listened carefully to the instructions, and did all my best to understand and follow them. In this specific task, I was instructed to perform quick flexion of my right wrist on my own initiative, at any time I felt the "urge" or desire to do so, in the meanwhile to watch the

"clock position," and to try to remember the clock time of the first awareness of the urge to move. This was a quite demanding task; however, I managed to do so after some practices. The content of the instructions must have been internalized as mental representation stored in my mind and brain (called "task set" or "mental set" in the literature). If I had not memorized these instructions, and purposefully and consciously made efforts to follow them, it would have been apparently bizarre for me to have my right wrist flexed several 100 times during that period.

Third, when I was performing the required acts in accordance with the instructions, I was fully awake and conscious, well aware of what I was doing. I was not in sleep, coma, anesthesia, or under hypnosis. I knew exactly what to do, and endeavored to make things happen as intended. Of course, I was not aware of all the brain activities and mental processes; nobody can be. But it would be preposterous to infer that the unconscious cerebral activities preceding my self-initiated bodily movements, such as the RPs recorded in Libet's experiment, were irrelevant to my consciousness and agency. If I were in sleep, coma, anesthesia, or under hypnosis, it would be likely that they would not have occurred. If I were not instructed to perform such acts, or just had failed to follow the instructions, it would be likely that they would have not been recorded with the patterns that Libet has found in his experiment (Keller and Heckhausen 1990). Therefore, it seems absurd to infer that my consciousness, agency, and intention to perform such acts as instructed play no role in bringing about the unconscious brain activities associated with the seemingly self-initiated bodily movements in the experiment.

The above observations naturally exclude the interpretation that some specific brain activities leading to my self-initiated bodily movements, such as the RPs, occurred unconsciously, because (1) I was awake and conscious all the way; (2) I was consciously executing my intention or plan to participate in the experiment; (3) the instructions represented and stored in my mind and brain played a crucial role in completing what I was required to do in the experiment; and (4) I made effort to follow the instructions to bring about the intended actions. Moreover, given the content of the instructions and the nature of the experimental design, I was required to perform quick flexion of my right wrist on my own initiative, at any time I felt the "urge" or desire to do so. These acts were brought about by me consciously and purposefully, rather than being fully spontaneous according to the standard interpretation. I simply would not have my right wrist flexed several 100 times in half an hour, had I not made efforts to follow the experiment instructions. In other words, to follow the instructions, I tried to make many times of the requested bodily movements happen during the period of experiment. So, the RPs were also made happen by me, who was consciously and intentionally making the related acts happen, even if the exact time of the RPs' onsets was unconscious to me. I knew that I needed to make the specific bodily movements happen many times during a short period, spontaneously as far as I could, without any preplan or at any regular rhythm. This experimentally structured behavior is not similar to my typical voluntary and spontaneous behavior in real life.

Klemm rightly points out that consciousness, especially human consciousness, raises the complexity of agency to a new level (A9 in his axioms list), and opposes

to the proposition that conscious mind is not a source of agency (P7 in his propositions list). It remains a long way to figure out the role of consciousness in the work of agency, and vice versa. A common pitfall to watch out for is the fallacy that may be called "consciousness localization." All mental states and processes are divided into two categories: conscious and unconscious. Consciousness is then attributed to some mental states and processes that the agent can be aware of. Thus, consciousness is a feature of an individual state or process that one can be conscious of. The nature of the fallacy is the neglect that consciousness is in the first place, a feature of the whole agent. It is not uncommon to find statements in the literature like this: Behavior x is caused by unconscious state or process y, or, behavior x is caused by conscious state or process y. What has been widely neglected in the above reasoning is the fact that most data obtained from experiments to support the statements are obtained while the subjects are awake and conscious, not in sleep, coma, anesthesia, or under hypnosis. The agent's being conscious may likely play an essential role in bringing about the behaviors that are recorded and measured. Unless the agent's whole consciousness has been fully eliminated, it is unwarranted to infer that an agent's behavior is solely caused by an individual mental or brain state or process, unconscious or not.

Understanding agency matters. Neuroscience is becoming more and more important in exploring the cognitive and neurobiological underpinnings of the work of agency. Klemm's comprehensive review has provided a state-of-the-art survey from the perspectives of neuroscience. This chapter attempts to supply a complementary, and more philosophical perspective, and to join the ongoing interdisciplinary effort to understand the nature of agency.

Acknowledgments This work was supported by grants from the Ministry of Education of China (13JDZ004) and from the Social Sciences Foundation of China (11&ZD187).

References

Bratman, M. E. (1987). *Intention, plans and practical reason*. Cambridge: Harvard University Press.

Bratman, M. E. (1999). *Faces of intention: selected essays on intention and agency*. Cambridge: Cambridge University Press.

Bratman, M. E. (2007). *Structures of agency: essays*. Oxford: Oxford University Press.

Keller, I., & Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts: Voluntary vs. involuntary control. *Electroencephalography and Clinical Neurophysiology, 76,* 351–361.

Klemm, W. R. (2015). Neurobiological perspectives on agency: Ten axioms and ten propositions. In C. Gruber, M. Clark, H. Klempe & J. Valsiner (Eds.), *Annals of theoretical psychology: Vol. 12. Constraints on agency: Explorations of theory in everyday life*. New York: Springer.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Science, 8,* 529–566.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain: A Journal of Neurology, 106*(3), 623–642.

Zhu, J. (2003). Reclaiming volition: An alternative interpretation of Libet's experiment. *Journal of Consciousness Studies, 10*(11), 61–77.

Chapter 6
A Grand Synthesis: Aided by Considering Systems 1 and 2 and Incentive Motivation

Frederick Toates

Klemm (Chap. 4, this volume) regrets that the concept of agency features rather little in neuroscience. However, expressed in somewhat different terms, it forms the topic of a number of studies, particularly in the context of conscious and unconscious determinants of behaviour (Pockett et al. 2006). If Klemm is claiming that the notion deserves to be at centre stage in behavioural science, I agree. Surely, goal-directed behaviour and the associated agency constitute one of the defining features of what it is to be a living system, at least, I would say, in the case of birds and mammals. Hence, I welcome warmly this contribution to the discussion.

I see two fundamental but all too often neglected principles underlying behavioural control: (1) its goal-directed nature and (2) its susceptibility to reinforcement. One might expect that these would have a pride of place in psychology and neuroscience, much as the genetic transmission of information and evolution by natural selection do in biology. However, this is not the case. Although the notion of goal direction is evident in theoretical psychology (Carver and Scheier 1990; Miller et al. 1960; Toates 2006), it hardly takes the place of a foundational principle.

Of course, Skinner put reinforcement on the map but unfortunately alienated many by his somewhat dogmatic and exclusive approach. However, the cognitive revolution tended to cast Skinner to one side rather than rising to the challenge of assimilating his ideas into a bigger and integrative psychology. (Toates 2009)

Any discussion of agency needs to consider also the weakness and lack of agency, exemplified by helplessness and depression.

6.1 What Is Agency?

Despite my enthusiasm for the target article, it seems that Klemm employs ‘agency’ somewhat inconsistently. At one point, he uses it to refer to ‘any animal action that arises out of its nervous system’. I would prefer to see it defined more narrowly than

F. Toates (✉)
Open University, Buckinghamshire, UK
e-mail: frederick.toates@open.ac.uk

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_6

this, though I would hate to be dogmatic. To me, 'agency' implies a goal and the organism's attempts to match an actual state of the world to this goal. In such terms, the rat has a goal representation and exploits flexible means to get to the goal, so is showing agency.

Humans, of course, unambiguously exhibit agency. However, in addition, they might be said to possess a 'sense of agency' when they can articulate their capacity to perform goal-directed behaviour and their willingness to take such action as to bring reality into alignment with their stated goals. The human showing learned helplessness and debilitating depression might be said to have largely lost both agency and this sense of agency. A possible rat model of agency would, as described by Klemm, involve suppression of 'generalized voluntary agency'. This would be mediated at least in part by reduced dopaminergic activity, which will be described later.

In my terms, there can be 'bottom-up drivers of agency', to use Klemm's expression, possibly provided that the incoming sensory signal does not trigger an automatic response that bypasses the intentional processes.

In such terms, in spite of their effective and adaptive behaviour, one would not see the jellyfish or ant to be displaying agency. Klemm writes 'A spinal reflex, for example, might not be considered an expression of agency because it arises from local circuitry and is not necessarily engaged with the whole nervous system'. I agree with this, yet elsewhere he states that a reflex is 'the simplest form of agency'. As an issue of personal taste, I would not include 'acts of agency' arising from invertebrate command neurons. If one can include jellyfish under the term 'agency', why not single-celled species lacking any nervous system, such as the paramecium? They show adaptive behaviour by altering their behaviour in accordance with changes in the environment. Why not trees and flowers? They exhibit changing reactions in response to, or even in anticipation of, such things as changes in season and light levels. Defined too broadly, it seems that agency means little more than the existence of a viable and adaptive organism. A species lacking a capacity for agency might seem to be almost synonymous with an extinct species.

Klemm's Axiom 8 states that 'Initiating agency, such as deciding what to do, typically requires assessment of anticipated positive and negative consequences and their reward value'. This apparently narrower definition would seem to me to be the essence of what is meant by agency. There is a cost–benefit assessment done. For example, rats appear to weigh up the relative cost–benefit of one choice as opposed to another, and disrupting dopamine neurotransmission alters the weighting of the options. (Salamone et al. 2007)

6.2 Other Bodies of Theory with Which to Integrate

I would suggest that the development of Klemm's account of agency might be enriched by a consideration of some theories that have traditionally stood somewhat distinct but are coming increasingly into convergence (Toates 2006, 2014). They are described in this section.

6.3 The Division into System 1 and System 2

A broad theoretical development has taken root in psychology and to this Klemm makes implicit reference: recognition of the coexistence of two fundamentally different systems of control (Carver et al. 2009; Epstein 1994; Evans 2008; Perner 2003; Toates 1998, 2006, 2014). It might be very helpful when pondering the nature of agency to look carefully at this distinction.

System 1 is fast, reactive to events in the world, seen widely across species, old in terms of evolution and early to appear in the development. The processing by System 1 is not available to conscious introspection, though the outputs are. By contrast, System 2 is slow, reflective, evolutionarily new and appears late in development. It facilitates the use of representations of events in the control of behaviour. Although System 2 is not a uniquely human possession, it is seen at its most elaborate in humans. In humans, processing within System 2 is open to conscious introspection. I would associate this system with goal-directed behaviour involving a consciously accessible representation of future states and moves towards them.

The weight of responsibility for the control of behaviour varies between System 1 and System 2 corresponding to a number of factors (see Toates 1998, 2006, 2014):

Experience With extensive experience of performing a task, control tends to shift from System 2 to System 1. Behaviour becomes more automatic and habitual.

Development System 1 matures faster than System 2. Hence, young animals place a greater responsibility for control on System 1, relative to System 2 (Steinberg 2008).

Chemicals A range of evidence suggests that low serotonergic activity gives a bias in favour of System 1 (Carver et al. 2009). For another example, alcohol creates 'alcohol myopia' (Steel and Josephs 1990), whereby weight is placed upon the immediate present situation and away from future.

Brain Damage For example, damage to the prefrontal cortex can compromise System 2 control, while leaving System 1 relatively intact. There can be an increased tendency to mimic the actions of others, the so-called utilization behaviour (Hurley 2006).

Species Compared to other species, humans appear to have placed the greatest weight upon System 2 controls, relative to System 1.

Although these are two distinct systems showing very different properties, they are dynamically interactive, such that many, if not all, instances of behaviour are controlled by both systems (Kiefer 2012; McBride et al. 2012). Anatomically, there are some shared brain processes. Outcomes produced by System 1 can trigger the engagement of System 2, particularly if they violate a goal set by System 2. This fits the notion discussed by Klemm that a (not *the*) function of consciousness is to monitor the 'consequences of unconscious agency'. System 2 can alter the sensitivity of System 1 in its reaction to stimuli such that reactions to stimuli accord with high-level goals (McBride et al. 2012; Toates 1998).

Sometimes the systems act in concert, whereas at other times it appears that they pull in opposite directions. Such conflict can take various forms (Perner 2003). Turning left for home automatically by habit in the face of a high-level goal of turning right for the supermarket is one such example. For another example of conflict, resisting temptation might be characterized as System 1 urging engagement in the tempting activity, whereas System 2 is exerting restraint based upon long-term considerations (Toates 2014). The opposite type of conflict would be exemplified by ingesting a foul-tasting medicine. System 1 might be triggering disgust and withdrawal, whereas System 2 would be urging engagement in the long-term interests of recovery of health.

System 1 would seem to have some of the characteristics of the kind of stimulus–response (S–R) processes much loved by the early generation of (non-Skinnerian) behaviourists. It is triggered into action by physical stimuli; in other words, the 'affordance' that they offer (McBride et al. 2012). However, it does not prescribe a fixed movement in response to a given stimulus. Rather, it specifies an action with some local negative feedback and thereby exhibits a degree of flexibility.

System 2, based upon internal representations of the world, either actually present or expected and desired, seems to fit high-level goal-directed cognitive principles. Hence, rather than two rival schools of psychology, we can see two different processes that co-exist within the same brain.

I wonder whether Klemm would attribute agency equally to Systems 1 and 2. I am tempted to associate it with System 2 but not with System 1. However, I am ambivalent.

6.4 The Principle of Incentive Motivation

Researchers in motivation are increasingly rejecting the earlier drive model, and instead finding an incentive motivation model better fits the evidence (Berridge 2001, 2004; Bindra 1978; Toates 1986). The fundamental assumption is that motivational states, such as hunger, thirst, sexual desire and drug seeking, are aroused by incentives, such as food, water, a partner and drugs. Incentives create the motivation to engage with them. The power of objects in the world to trigger engagement is also exemplified in certain brain-damaged people by utilization behaviour, where the person might suddenly and inappropriately grab someone else's apple and start to eat it. (McBride et al. 2012)

In the absence of the incentive, the appropriate motivational state and goal-directed activity can be aroused by cognitive representations of incentive objects (Kavanagh et al. 2005). Functional neuroimaging points to some of the same brain regions being activated by triggering memories of incentive objects as compared to that triggered by the actual object. An implicit assumption within motivation theory is that the strength of motivation increases as the distance between the animal and the object decreases.

6.5 The Role of Dopamine and Serotonin

Dopamine appears to act at different levels (Berridge 2001), which might map onto Systems 1 and 2, just described, as well as being central to incentive motivation. Acting at a sub-cortical level in the pathway from the ventral tegmental area to the *nucleus accumbens,* dopamine has an energizing effect upon behaviour (Alcaro et al. 2007). As a result of dopaminergic activation, incentives in the world acquire a kind of magnetic pull, showing some properties of System 1 control. Acting at a cortical level, dopamine appears to facilitate truly goal-directed behaviour, i.e. a System 2 control, in humans characterized by conscious intentions.

Boosting dopamine levels has the effect of giving incentive salience to objects in the world and increasing engagement with them, surely a cardinal feature of agency. This is observed in the case of encephalitis lethargica (sleeping sickness; Sacks 1976) and in the manic phase of bipolar disorder (Cousins et al. 2009). Giving precursors of dopamine as a treatment for Parkinson's disease is sometimes followed by addiction. (reviewed by Toates 2014)

Active but restrained ('controlled') engagement in the world is associated with System 2 activity, mediated via high levels of dopaminergic and serotonergic activity (Carver et al. 2009). It appears that impulsiveness, mediated via System 1, is based upon relatively high dopamine levels, accompanied by relatively low serotonin levels. Conversely, depression and helplessness, also mediated, it is argued, by System 1, appear to be associated with relatively low levels of both serotonergic and dopaminergic activity.

6.6 Conscious and Unconscious Processes

Basics

I agree entirely with Klemm when he states that 'conscious operations are explicitly deliberative, analytical, focused on the novel and unlearned, and relatively slow. Unconscious operations seem to operate at high-speed on well-learned memories'. This appears to map very well onto System 2 and System 1, respectively. So much so that one might wonder why use the terminology of 'Systems', when 'unconscious' and 'conscious' would seem to do just as well and have greater intuitive appeal. Use of the term 'System' facilitates cross-species comparison between humans and non-human species, where for the latter we cannot really use the conscious–unconscious dichotomy.

6.7 Does Consciousness Have Behavioural Efficacy?

Klemm discusses the thorny issue of whether consciousness has any causal efficacy, noting in Axiom 9 that 'A growing fad in neuroscience is a belief that consciousness cannot do anything'. The argument for and against seems to me to often miss a central point, as follows.

A key issue here concerns the relationship between brain events and conscious mental events. Klemm seems implicitly to favour a version of identity theory. So, suppose that every conscious state (e.g. C1) is observed to be invariably accompanied by a unique pattern of neural activity (N1), whereas a different conscious state (C2) is invariably accompanied by a different pattern of neural activity (N2). We know that altering the activity of the brain, by, for example, electrical stimulation alters conscious experience, so the temptation is to say that the brain state *causes* the conscious state. However, a strict identity theory would see the brain state, not as causing the conscious state, but as being its correlate.

For an analogy, does the flow of sodium and potassium ions *cause* the action potential?[1] According to identity theory, presumably it does not, for there is not a time sequence of, first, influx of sodium and then the action potential. Rather the action potential *is* the flow of ions, expressed in other words and both are caused by, for example, the occupation of receptors by neurotransmitter.

Those *brain processes* that are associated with conscious awareness are clearly of causal efficacy. For example, surely it would be foolish to deny that the combination of the pattern of neural activity (N1) triggered in the brain by, say, a thorn stuck in the back and the aversive conscious correlate (C1) are involved in the agency of asking someone for help. No automatic defensive reflex could produce such behaviour. It appears that the issue really amounts to the following: Could one remove C1, leaving N1 intact and still be expected to make the same request for help? This might prove to be rather like asking whether one could remove the ionic flows, while still leaving the action potential.

It could simply be inevitable that conscious states appear when a certain level of brain complexity emerges in evolution. That is to say, the brain with the combination (C1N1, C2N2, C3N3, etc.) has proven advantageous as compared to a less complicated system without such a correlated consciousness (N7, N8, N9, etc.).

Based upon the philosophical principle of identity theory, my hunch is that it will forever be impossible both empirically and conceptually to decide whether conscious states are of causal efficacy, i.e. whether N7, N8, N9, etc. can do the job as well as C1N1, C2N2, C3N3, etc.

Viewed in such terms, some of the discussions around causal efficacy need careful qualification. For example, in 'New Age' literature, the existence of the placebo effect is sometimes used to argue for the efficacy of consciousness. It is said that a mere expectation can have effects upon the body. However, such a 'mere expectation' is itself embodied in neural activity, the result of information that enters the brain through neural activity arising in the ears and eyes. Any psychosomatic effect can be framed in exactly the same language.

We know that the world of fantasy, operating independently of current sensory information, is associated with the activation of some of the same brain regions as underlie actual action (Tian and Poeppel 2012) and can be of causal efficacy. For example, while a changing level of illumination is the normal trigger to changes in pupil size, simply *imagining* different levels of illumination also changes pupil size

[1] I am indebted to Steven Rose for this analogy.

in a similar direction to changes in the light stimulation (Laeng and Sulutvedt 2014). Some women can trigger orgasm by thought alone (Whipple et al. 1992)! Practising skills in the imagination improves actual performance at the skill (Baumeister et al. 2011). Yet, the implicit assumption of identity theory is that, for each such event in the conscious theatre of fantasy, there is a corresponding pattern of neural activity.

Klemm writes: 'It seems natural to suggest that main function of consciousness is to refine, amplify, and perfect unconscious processes'. Yes, but I would express this as the main function being that of setting intentions at the top of a hierarchy within System 2 and coordinating conscious and unconscious processes to meet this conscious goal.

6.8 Willed Actions and Free Will

Klemm notes that '…free will is possible in the sense that one can choose among multiple options and even choose against options that provide the most positive reinforcement at the time'

Indeed and we tend to attribute virtue in proportion to the extent to which a person follows the call of long-term and altruistic goals as opposed to short-term hedonic rewards. Of course, Eve was condemned for getting this calculation wrong in the Garden of Eden (Toates 2009, 2014). However, this capacity is only a necessary condition for the existence of free will but not a sufficient one, since the long-term altruistic goals could just as easily be determined as the short-term hedonistic ones.

In fact, I really do not know what the term 'free will' means though I think that I know what it is *not*. If one could predict behaviour with absolute certainty based upon genes, the initial internal environment of the zygote and all that happens from that point on, one might see this as undermining the existence of free will. This is, of course, quite impossible to do in practice.

Possibly in order to act freely, the conscious mind would need to have access to a random number generator that threw up representations of a range of possible actions and expected outcomes of them. These would need to be free from past influences. However, it is hard to see how the process of *selection* from amongst these options could be free from determinism by the past.

6.9 Relevance of System 1 and System 2

Klemm writes: 'Decisions can be made unconsciously or consciously. In the former case, those willed actions that are simple and habitual need no conscious support'. Here, we need to qualify what is meant by 'willed actions'. If behaviour is captured by a powerful incentive stimulus acting through System 1, possibly against the individual's conscious intentions, can it really be described as 'willed'?

Klemm writes: 'This is a "bottom up" idea, where consciousness (at the top) is a recipient of input that has gained a competitive advantage for conscious "viewing". But why couldn't conscious sense of self be a "top down" process, where the

conscious mind "looks in" on local processes and selects which ones it was to pay attention to and modify?' This is exactly as I see conscious agency acting as System 2.

I agree with Klemm's list of reasons why Libet's experiment does not really undermine either (1) conscious agency or (2) the notion of free will (whatever the latter might mean!). Indeed, I presented a similar argument some years ago (Toates 2006). Slow and reflective conscious control can delegate responsibility for simple actions, such as a button press, to a lower hierarchical level. One assumes that Libet's participants made a conscious choice in the first place (had a 'pre-existing intent' in Klemm's terms) on whether to participate or not and subsequently when in the laboratory to give 'permission' to lower controls to act. Slow and reflective conscious controls could act at the level of sensitizing local neural circuits. As Klemm notes, in monkeys, priming cues sensitize subsequent responses.

Klemm writes: 'Common experience teaches that agency does not occur in unconscious states such as sleep, coma, or anaesthesia'. To split hairs, some agency might occur in some such states. Sufferers from night terrors take limited evasive action such to move away from the threat 'in the bed' and scream (Cartwright 2004). Does this constitute agency?

This is discussed further in the next section.

6.10 States of Brain Damage

Klemm writes 'Then there is the matter of wakefulness, which in humans is typically associated with consciousness'. Consider the following:

Our intuitive sense of awareness in another individual is based upon their reaction to prompts, such as 'If you can hear me, please raise one finger'. However, with the help of modern medicine, an increasingly large percentage of people with serious brain damage are being kept alive. Although they show clear signs of wakefulness and sleep–wake cycles, they show only reflex responses and no purposive overt behaviour. They might not respond overtly to prompts, but does this mean that they are lacking awareness and all forms of agency?

Could it be that people in so-called vegetative states have at least a residual level of conscious awareness and agency in some sense but are simply unable to perform the necessary motor acts of responding? This question was addressed by Adrian Owen and colleagues in pioneering research employing functional neuroimaging (Owen 2013).

There are two basic foundations of the approach of these researchers:

Functional magnetic resonance imaging (fMRI) can be used to detect neural correlates of conscious awareness while the observed patient appears to be making voluntary choices.

To imagine doing a task is associated with activation of some of the same brain regions as are involved in actually performing the task. For example, to imagine a motor task, such as squeezing the hand, involves activation in the motor and premotor cortex. By contrast, to imagine navigating a house involves activation of the parahippocampal cortex.

In control participants, such patterns of brain activation are indicative of understanding the question and voluntarily choosing to act upon it. Hence, Owen and colleagues argue that such brain responses can serve as proxies for the corresponding motor actions. Could brain-damaged patients show similar patterns of activation that serve as proxies for the motor actions that they are unable to perform? If so, this might indicate retention of at least some degree of conscious awareness and capacity for voluntary choice. The results showed that a considerable proportion of brain-damaged patients in so-called vegetative states did exhibit patterns of brain activity that could be used as proxies for the overt motor responses.

Sceptics might argue that priming with a keyword like 'squeeze' triggers activity in motor areas of the brain, corresponding to unconscious (i.e. System 1) processing. Owen (2013) counters this by noting that appropriate activation of motor regions was observed for the full 30-s period following each scenario prompt. Such extensive activation is not seen in response to single words. Where any single-word triggered activation is observed, it is of very short duration compared to the 30 s seen here in response to questions. It would seem then that the patients showed goal-directed activity that was sustained in the absence of any further cuing over a 30-s period. The question took 'possession of the mind', to use a popular expression (Dehaene and Changeux 2011).

When the same questions are set to people under anaesthesia, there is not such a discrimination of brain activity patterns. Owen (2013, p. 122) writes:

> …healthy volunteers who are measurably nonaware (i.e. unconscious) are not able to generate the characteristic pattern of brain activity that is associated with imagining playing tennis, suggesting that awareness is likely to be necessary for this response to occur in patients.

In response to biographical questions, the researchers were able to get correct 'yes' or 'no' responses from a 'vegetative' patient by asking him to imagine a motor task if the answer is yes or a navigation task if the answer is no. This performance involves a number of processes that are normally associated with conscious awareness (Owen 2013):

- *Long-term memory*, in order to assess the meaning of the words of instruction and associate them with the corresponding action
- *Working memory*, in order to hold the target online during the period until the following question
- Attentional switching from one situation to another
- Response selection

6.11 A Ghost in the Machine?

Let me now throw a spanner in the works, though it is one that might at least cast light on the conceptual impasse posed by identity theory and the question of the agency of conscious states. Klemm writes: 'We know that consciousness comes from and resides in the brain'. For most of my colleagues, I have little doubt that

this would be a viable candidate to become number 11 in Klemm's list of axioms. I might agree with them on every second day of the week. Otherwise, I would see it as no more than a proposition. How do we know with such confidence where consciousness resides? Forgive me for playing devil's advocate but at times we need to question even our most basic assumptions.

Without doubt, the state of the conscious mind, as reported by introspection, is closely associated with the activity of the physical brain. So much can be demonstrated experimentally. But is the conscious mind something that literally *resides* in the brain? Does it exist in space at all? Of course, we all employ expressions like 'the article I have just written is *in* my computer', which usefully and uncontroversially distinguishes its hardware base. So, I hope that my point is more than irritating hair splitting.

Presumably, the assertion that consciousness is in the brain is a short hand for saying that information can only get into the brain by means of transduction into electrical signals in neurons. It can only get out by transduction from efferent neural signals.

I am very cautiously looking over my shoulder as I write these words. Only at an advanced stage in my career would I feel able to do so and thereby risk the accusation that I have gone soft in the head. However, I have read sufficient heretical works on such phenomena as near death experiences (van Lommel 2010) and remote viewing (Targ 2012) by serious scientists and clinicians, as well as the general critique of the notion of consciousness residing in the brain (Sheldrake 2012) to feel that we should at least seriously question the safety of this entrenched assumption.

After all, some of the pioneers of psychology (e.g. William James, Pierre Janet) and neuroscience (e.g. Hans Berger, John Eccles, Wilder Penfield, Charles Sherrington) held 'paranormal views' that would be rejected without question as foolish heresy by modern neuroscience. Donald Hebb took telepathy seriously (Hebb 1951). Yet I am hard put to discover the evidence which has arisen since the time of these greats, so as to implicitly or explicitly dismiss their arguments and assert with such confidence that consciousness resides in the brain.

6.12 Violence, Responsibility and Free Will

Klemm writes 'But agency is a very practical matter in the real world of human behaviour'. Indeed it is and the related issues of determinism and free will, including the aspects described earlier here, perhaps come into their sharpest focus and with the most controversy and profound social implications in the context of violent crime.

For example in the case of killers, the conclusion to such reflection can be quite literally a matter of life or death: the gas chamber or a hospital for the criminally insane. The legal profession watches developments in neuroscience with keen interest (Kaplan 2006) and their relevance will most likely increase in the coming years. Defence attorneys doubtless scrutinize any new developments that might provide evidence of brain abnormalities in their clients. I would suggest that the notion of agency and the distinction between System 1 and 2 could yield some useful insights.

6.13 Where Innocence Is Assumed

Consider a scenario enacted in 1987 (Denno 2003; Levy and Bayne 2004). A 23-year-old Canadian, Kenneth Parks, drove 14 miles across town, where he proceeded to kill his mother-in-law and almost to kill his father-in-law. Parks admitted to these actions but was acquitted of all charges, in that they were described as 'involuntary'. The defence case was that Parks, a sufferer from sleepwalking (somnambulism), was not conscious at the time of the killing. Shortly after the assault, Parks turned himself in at the local police station. He had no history of violence and was said to have got on well with his parents-in-law.

So, was Parks exhibiting agency? At some level in the brain, he must surely have been exhibiting goal-directed behaviour in the sense of the hierarchical organization of motor output, since the killing was a *novel* action. To produce novel actions is sometimes seen as one of the essential defining roles of *conscious* processing. It has been noted that, in other instances of sleepwalking, an attempt to interrupt the behaviour can be met with aggression (Cartwright 2004), suggestive of the thwarting of a goal-directed action.

What made Park's case different then from murder? Presumably, the assumption might be somewhat along the following lines. The assault on the parents-in-law was brought about by the sensitization of hierarchical structures by a series of linked goals (e.g. get to the house, gain entry, find victims, etc.). However, no such goal involved the top-down representation of the parents-in-laws' death with associated positive effect on both anticipating and subsequently achieving this goal. That is to say, there was no 'prior intent' (Pressman 2007); Parks would not have made a prior conscious cost–benefit analysis of the strategy of going and killing them. To adapt a criterion used to describe local automatisms, such as the alien hand syndrome, Park's behaviour formed 'part of no larger plan of action' (Levy and Bayne 2004, p. 211).

Sleepwalking appears to be associated with a reduction in the activity of frontal regions of the brain (Pressman 2007). As Pressman notes, damage to the frontal lobes is associated with heightened tendencies to aggression. He suggests that, in sleepwalking, limbic structures underlying aggression could be disinhibited with, in a very small minority of sleepwalkers, the acts of violence. Furthermore, where violence is involved, this is commonly preceded by an extensive period of stress, which could further act to shift the balance towards limbic structures (from System 2 to System 1).

Next, consider someone to be in a state of 'high emotional charge' as a result of, for example, suddenly finding themselves abandoned by a spouse. Totally out of character, the individual impulsively stabs the spouse and then claims no memory of having done so (McSherry 1998). Dissociation is frequently used in their defence, a state sometimes made more likely by excessive alcohol intake and stress. The behaviour is said to be an 'automatism'. In the terms of the present argument, this would amount to a dissociation between System 1 and 2, with System 1 usurping control from System 2.

6.14 Where Guilt Is Attributed

Consider next two tragic cases from England where one might argue that System 1 dominated but no such mitigating circumstances were permitted, in contrast to the cases just described.

The first concerns one of the country's most notorious cases: that of Ian Huntley, who was found guilty of the murder of two schoolgirls in 2002 (Yates 2005). It appears that Huntley lured them into his home in Soham on the pretext that his girlfriend, a teaching assistant at their school, was at home and she would be very pleased to meet them. Once inside, the most likely scenario is that Huntley made sexual advances towards them, which were rejected. At this point, it appears that he flew into a rage and strangled them. One can speculate that they might have said something along the lines of 'We will tell our parents about you'. An attempt to avoid this outcome could have played a role, in which case there would have been some intentional component to the action. Whatever the actual events, it seems most likely that the violent action was triggered impulsively with little or nothing in the way of prior planning.

Another case concerns the murder of a 25-year-old woman in her Bristol apartment in 2010 by her neighbour, Vincent Tabak.[2] It seems very doubtful that this was planned. Tabak, who held a doctorate in human systems behaviour, would surely have known that in such cases suspicion falls first upon relatives and neighbours. He admitted to the killing but claimed that it was accidental, caused by putting his hands over the victim's mouth to stop her screaming. The jury did not believe this story and he was found guilty of murder. Police investigations revealed that Tabak had an extensive prior history of viewing violent pornography, involving strangulation. One might reasonably speculate that such viewing and extensive use of corresponding fantasy had sensitized his motivational processes of sex-linked aggression. I would suggest that, in a highly aroused state and in the presence of a suitable target, his System 1 controls suddenly dominated, forcing out any potential restraining factors.

So, what is it about the last two cases that make them different from that of, say, killing a spouse who has just announced his exit from a marriage or killing during sleep? Why is the law likely to look very differently upon such cases? Presumably, it is the context. The last two had no justified reason to be in the situation in which the killing took place. Their long-term histories were ones of behaviour judged by society as unacceptable and potentially leading to disaster, as was their goal-directed actions immediately prior to the killing.

A legal implication of similar considerations was raised at the trial of the Chicago serial killer, John Wayne Gacy. A psychiatrist observed that the killer would excavate graves for future victims, adding that (Sullivan and Maiken 1983, p. 346):

[2] There appears to be no book describing this case. However, at the time of writing there are a number of serious websites active, which describe the details: http://www.theguardian.com/uk/2011/oct/28/vincent-tabak-porn-searches-jury

http://en.wikipedia.org/wiki/Murder_of_Joanna_Yeates.

> I don't think that a person who *plans* to have an irresistible impulse in the future could be considered *having* irresistible impulses.

Gacy could hardly have pleaded surprise on finding himself in each situation of killing, since there were 33 of them and they were carefully planned. However, one might argue that Gacy had irresistible impulses *when in the presence of his victims* but not at the planning stage.

Gacy exemplifies what is perhaps both the most feared and the most extensively documented of such crimes: that of the serial killer who murders for sexual gratification (Toates 2014). It would seem that most such killers came from a childhood background of extreme rejection, violence and bullying (Miller 2014), a common feature being the feeling of a very powerful grudge against society. The early emergence of sexual desire appeared on the scene at a time of the experience of such intense negative emotion and it would seem that, by means of a process of sensory preconditioning, there was a sex–violence fusion (MacCulloch et al. 2000). Neurons that fire together wire together.

While of course not condoning the crime, where such an aversive background can be identified, liberal opinion tends to see it as something of a mitigating factor that might mean an escape from the gas chamber. Are some people, because of their development, less free than others? I really do not know. It could be argued that as a result of their early experiences and developmental trajectory, they (1) have high levels of anger and (2) are lacking a process of empathy (Miller 2014), and are given an irresistible bias towards such behaviour.

Although it would seem that most such people had profound injustice done to them, in other cases, biographers are hard put to find anything that seems at all commensurate with the brutality of the crime (Miller 2014). Examples include some of the most notorious, such as Jeffrey Dahmer (USA), who suffered mainly from isolation and neglect (Nichols 2006), and Ian Brady (England), who was angered by being born working class and illegitimate (Lee 2012). Maybe in such cases, the process of mental rumination on a perceived injustice got into a self-reinforcing positive feedback loop such as to create a deviation from healthy development comparable to that of those having experienced extreme abuse.

The crimes of serial killers appear to be unambiguously under the control of System 2. Indeed, they would need to be in order to evade capture for so long. These individuals are described as experts in meticulous goal-directed planning (Miller 2014). Opportunities are weighed up and cost–benefit analyses of prospective actions performed. This would meet the criterion of guilt set by the American legal system that the individual 'had the mental capacity to have committed the act consciously, knowingly, and purposively' (Miller 2014, p. 20). Ted Bundy (USA) is a textbook example of meeting such criteria, where considerable foresight and planning were evident, including crossing state borders to evade detection and use of a plaster cast to simulate a fractured arm and thereby solicit help from the intended victims. (Rule 2006)

6.15 Conclusion and Discussion

I would suggest that further progress on understanding agency requires a tighter and clearer definition of exactly what is meant by the term. On balance, I think that I would not include jellyfish and ants in an account of agency since this might detract from the particular processes that are evident in mammals. However, it could prove valuable to compare and contrast across species concerning how behaviour is instigated, rather as how vertebrates and invertebrates have found very different solutions to the same problem of transducing light into a signal.

A closer integration with the bodies of theory and data associated with the notions of Systems 1 and 2, as well as incentive motivation and the role of neurotransmitters, could also prove useful. I have no doubt that dopamine plays an important role in the phenomenon of 'initiative' and individual differences in this, as described by Klemm (see Previc 2011). Also, dopamine is implicated in the trade-off between immediate and long-term reward, discussed by Klemm.

Notions of human agency are of central relevance to some conditions that are literally matters of life and death, such as the responsibility of killers for their actions and decisions on whether to terminate the lives of people in so-called vegetative states (Fernández-Espejo and Owen 2013). The ethical implications of showing the kind of intentionality implied by a 'yes' or 'no' to the question 'Shall we switch off your life support?' are profound indeed.

The law already appears to make an implicit assumption that accords with thinking in terms of System 1 and 2 and the developmental shift of weight towards System 2. Thus, children are not subject to the same punishment as adults (e.g. in the USA not sentenced to death) on the grounds that their immaturity precludes them being able to form the necessary ethical distinction between right and wrong. I would suggest that in some cases they might be able to reason right and wrong as well as an adult but under are unable to exert a comparable degree of agency based upon it.

What is it about agency that means that Kenneth Parks walked away a free man, whereas Ian Huntley and Vincent Tabak languish in jail? It might not hinge on the operation of System 1, which seems not to distinguish these cases. Rather, I would expect legal arguments to centre around a consideration of whether System 1 was sensitized top-down by System 2 at the time of the crime or whether the crime was done in spite of a restraining influence of System 2. In Klemm's terms, it might be argued that in no case was there a murderous 'prior intent' and they fit Klemm's statement that 'Other decisions can be triggered as a complex reflex by external compulsion without prior intent'.

Rather the implicit logic that distinguishes sleepwalking from the other two seems to be based on the context in which System 1 functioned. Insofar as Tabak was free to do anything, his freedom was expressed over the long term in the choice of watching violent pornography and exercising his associated fantasy. To use Klemm's argument, agency was 'smeared out over time'. However, even here, he might have been subject to influences quite outside his conscious experience. Thus, aggression can be primed by role models and imitation in the face of a denial by the individual that they are having any effect (Hurley 2006).

References

Alcaro, A., Huber, R., & Panksepp, J. (2007). Behavioral functions of the mesolimbic dopaminergic system: An affective neuroethological perspective. *Brain Research Reviews, 56*(2), 283–321. doi:10.1016/j.brainresrev.2007.07.014.

Assadi, S. M., Yucel, M., & Pantelis, C. (2009). Dopamine modulates neural networks involved in effort-based decision-making. *Neuroscience and Biobehavioral Reviews, 33*(3), 383–393. doi:10.1016/j.neubiorev.2008.10.010.

Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. (2011). Do conscious thoughts cause behavior? In S. T. Fiske, D. L. Schacter, & S. E. Taylor (Eds.), *Annual review of psychology* (Vol. 62, pp. 331–361). Palo Alto: Annual Reviews.

Berridge, K. C. (2001). Reward learning: Reinforcement, incentives, and expectations. *Psychology of Learning and Motivation: Advances in Research and Theory, 40,* 223–278.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology and Behavior, 81*(2), 179–209. doi:10.1016/j.physbeh.2004.02.004.

Bindra, D. (1978). How adaptive-behavior is produced—perceptual-motivational alternative to response-reinforcement. *Behavioral and Brain Sciences, 1*(1), 41–52.

Cartwright, R. (2004). Sleepwalking violence: A sleep disorder, a legal dilemma, and a psychological challenge. *American Journal of Psychiatry, 161*(7), 1149–1158. doi:10.1176/appi.ajp.161.7.1149.

Carver, C. S., Johnson, S. L., & Joormann, J. (2009). Two-Mode models of self-regulation as a tool for conceptualizing effects of the serotonin system in normal behavior and diverse disorders. *Current Directions in Psychological Science, 18,* 195–199.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97*(1), 19–35.

Cousins, D. A., Butts, K., & Young, A. H. (2009). The role of dopamine in bipolar disorder. *Bipolar Disorders, 11,* 787–806.

Davis, D. (1991). *The Jeffrey Dahmer story: An American nightmare*. New York: St Martin's Press.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200–227. doi:10.1016/j.neuron.2011.03.018.

Denno, D. W. (2003). A mind to blame: New views on involuntary acts. *Behavioral Sciences and the Law, 21*(5), 601–618. doi:10.1002/bsl.553.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59,* 255–278.

Fernandez-Espejo, D., & Owen, A. M. (2013). Detecting awareness after severe brain injury. *Nature Reviews Neuroscience, 14*(11), 801–809. doi:10.1038/nrn3608.

Hebb, D. O. (1951). The role of neurological ideas in psychology. *Journal of Personality, 20*(1), 39–55.

Hurley, S. (2006) By-passing conscious control: Media violence, unconscious imitation, and freedom of speech. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior*? (pp. 301–337). Cambridge: The MIT Press.

Kaplan, L. V. (2006) Truth and/or consequences: Neuroscience and criminal responsibility. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior*? (pp. 277–299). Cambridge: The MIT Press.

Kavanagh, D. J., Andrade, J., & May, J. (2005). Imaginary relish and exquisite torture: The elaborated intrusion theory of desire. *Psychological Review, 112,* 446–467.

Kiefer, M. (2012). Executive control over unconscious cognition: Attentional sensitization of unconscious information processing. *Frontiers in Human Neuroscience, 6*. doi:10.3389/fnhum.2012.00061.

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science, 25,* 188–197.

Lee, C.A. (2012). *One of Your Own*. Edinburgh: Mainstream Publishing.
Levy, N., & Bayne, T. (2004). Doing without deliberation: Automatism, automaticity, and moral accountability. *International Review of Psychiatry, 16*(3), 209–215. doi:10.1080/09540260400003909.
MacCulloch, M., Gray, N., & Watt, A. (2000). Brittain's Sadistic Murderer Syndrome reconsidered: An associative account of the aetiology of sadistic sexual fantasy. *Journal of Forensic Psychiatry, 11*(2), 401.
McBride, J., Boy, F., Husain, M., & Sumner, P. (2012). Automatic motor activation in the executive control of action. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00082.
McSherry, B. (1998). Getting away with murder? Dissociative states and criminal responsibility. *International Journal of Law and Psychiatry, 21*(2), 163–176. doi:10.1016/s0160-2527(98)00006-5.
Miller, L. (2014). Serial killers: II. Development, dynamics, and forensics. *Aggression and Violent Behavior*, *19*(1), 12–22. doi: http://dx.doi.org/10.1016/j.avb.2013.11.003.
Miller, G. A., Galanter, E., & Pribram, K. A. (1960). *Plans and the structure of behavior*. New York: Holt, Rhinehart, & Winston.
Nichols, D. J. (2006). Tell me a story: MMPI responses and personal biography in the case of a serial killer. *Journal of Personality Assessment, 86,* 242–262.
Owen, A. M. (2013). Detecting consciousness: A unique role for neuroimaging. *Annual Review of Psychology, 64,* 109–133.
Perner, J. (2003) Dual control and the causal theory of action: The case of non-intentional action. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 218–243). Oxford: Oxford University Press.
Pockett, S., Banks, W. P., & Gallagher, S. (Eds.). (2006) *Does consciousness cause behavior*? Cambridge: The MIT Press.
Pressman, M. R. (2007). Disorders of arousal from sleep and violent behavior: The role of physical contact and proximity. *Sleep, 30*(8), 1039–1047.
Previc, F. (2011). *The dopaminergic mind in human evolution and history*. Cambridge: Cambridge University Press.
Rule, A. (2006). *The stranger beside me*. London: Sphere Books.
Sacks, O. (1976). *Awakenings*. New York: Vintage Books.
Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology, 191*(3), 461–482. doi:10.1007/s00213-006-0668-9.
Sheldrake, R. (2012). *The science delusion: Freeing the spirit of enquiry*. London: Coronet.
Steele, C. M., & Josephs, R. A. (1990). Alcohol myopia—its prized and dangerous effects. *American Psychologist, 45*(8), 921–933. doi:10.1037/0003-066x.45.8.921.
Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28,* 78–106.
Sullivan, T., & Maiken, P. T. (1983). *Killer clown: The John Wayne Gacy murders*. New York: Pinnacle Books.
Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00314.
Targ, R. (2012) *The reality of ESP: A physicist's proof of psychic abilities*. Wheaton: Quest Books.
Toates, F. (1986). *Motivational systems*. Cambridge: Cambridge University.
Toates, F. (1998). The interaction of cognitive and stimulus-response processes in the control of behaviour. *Neuroscience and Biobehavioural Reviews, 22,* 59–83.
Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition, 15*(1), 75–118.
Toates, F. (2009). *Burrhus F. Skinner*. Basingstoke: Palgrave Macmillan.
Toates, F. (2014). *How sexual desire works: The enigmatic urge*. Cambridge: Cambridge University Press.

van Lommel, P. (2010) *Consciousness beyond life: The science of the near-death experience*. New York: Harper-Collins.
Whipple, B., Ogden, G., & Komisaruk, B. R. (1992). Physiological correlates of imagery-induced orgasm in women. *Archives of Sexual Behavior, 21,* 121–33.
Yates, N. (2005). *Beyond evil*. London: John Blake.

Chapter 7
The Neuroscience of Agency and Free Will

Markus E. Schlosser

Before I turn to my comments on Klemm's (Chap. 4, in this volume) target chapter, let me provide some notes about my background. I am a philosopher, and my main area of research is the philosophy of action. I started to read empirical research on human agency about four years ago in connection with taking up a research fellowship in a project on the philosophical implications of empirical studies of moral agency.[1] I was looking, first and foremost, for psychological and neuroscientific research on the various notions that lie at the core of philosophical theorizing about agency and free will—in particular, I was looking for research on intentional action, acting for reasons, decision making, long-term planning, free choice, and free action. I was struck by a number of things. First of all, although it is often remarked that neuroscience, and in particular cognitive neuroscience, is still in its infancy, I found an enormous amount of neuroscientific research on human behavior. Very little of it, however, was directly about the mentioned notions that are at the heart of philosophical accounts. One reason for this, no doubt, is that much of this research concerns the mechanisms and the details of movement control. Another reason, it seems, is that scientists tend to work with different concepts and conceptual frameworks. For instance, much of the literature that is relevant here can be found in the large body of research on executive control. The notion of executive control, however, is almost completely absent from the philosophical debate and from philosophical theorizing about agency. Further, it was surprisingly difficult to find overarching theories, overviews, and reviews that aim to integrate and unify the findings from different strands of research and different experimental paradigms. It was difficult, in other words, to see the big picture. Given this, I can only agree with Klemm, when he says that more neuroscience is needed (p. 1 in MS), and I would

[1] This project was entitled "Morality Beyond Illusions: Reassessing the Philosophical Implications of Empirical Studies of Moral Agency." It was funded by the Netherlands Organization for Scientific Research (NWO), led by Pauline Kleingeld, and hosted by the University of Leiden.

M. E. Schlosser (✉)
University College Dublin, Dublin, Ireland
e-mail: markus.schlosser@ucd.ie

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_7

add that overarching and integrative theories are needed in particular. In the more recent literature, one can find more attempts to integrate the findings from different strands of research, and Klemm's chapter is a welcome addition to this trend. In addition to the research mentioned by Klemm, one can find interesting theories and helpful overviews in Haggard (2008), Desmurget and Sirigu (2009), Cisek and Kalaska (2010), Gallivan et al. (2011), Momennejad and Haynes (2012), Desmurget (2013), and Brass et al. (2013).

Klemm provides his overview to the neuroscience of agency by way of ten axioms and ten propositions (at the request of the editor, as I understand). I found this to be an interesting and fruitful approach. I learned a great deal and I found myself in agreement with most of Klemm's claims.

There are, nevertheless, many issues in this rich chapter that deserve discussion. I decided to restrict my commentary to four main topics (in the following four sections). Most of my comments concern the particular claims and suggestions one can find in Klemm's contribution. But I will also use the opportunity to offer some more general reflections on the conceptual and methodological issues that arise for the scientific study of agency and free will. I will argue, in particular, that most neuroscientific studies of free will (and voluntary action) are based on an operational definition that is deeply flawed. With this, I hope to show that the philosophical concern with conceptual analysis and plausibility is not mere idle reflection if it concerns operational definitions that underlie experimental paradigms. Let me stress, though, that this criticism should also be understood as a plea for more interdisciplinary interaction. Neuroscience has delivered fascinating findings about agency, which should be of interest to philosophers. Experimental findings, however, are only as good as the experimental designs that deliver them, and experimental designs are only as good as the operational definitions and conceptual frameworks that underlie them. Given this, neuroscientists may well have something to gain by considering philosophical accounts of agency and free will, which are, after all, based on centuries of philosophical reflection and debate on the concepts in question.

7.1 Defining Agency

In his chapter, Klemm offers various reflections on the concept of agency and several characterizations of the nature of agency. The chapter does not provide a definite definition of agency, and it is not clear whether Klemm takes any one of the given characterizations as central. This is not a problem, I think, as one can distinguish plausibly between different *kinds* of agency, and as one need not give conditions that unify all kinds of agency in order to say interesting things about certain kinds of agency. In this section, I will first offer some comments on Klemm's characterizations of agency. Then I will outline how action and agency are usually conceptualized within philosophy. This will provide the background for a comment on Klemm's account of the process and sequence of intentional and conscious agency.

In the first section of Klemm's chapter, one can find the following suggestions on how to think about agency: agency as acting in the world, agency as animal action that arises out of its nervous system, and agency as something that is applicable to the whole animal nervous system (pp. 1–2 in MS). All three suggestions sound plausible, but there are also a number of issues. First, any definition or characterization of agency in terms of *action* is unsatisfactory as long as we are not given a further definition or characterization of the nature of action. Klemm, it seems, takes it for granted that we know what action is—that we know what distinguishes action from mere movement or unintentional behavior (we will return to this below). Given this, the first characterization is uninformative, if not circular. The second characterization also presupposes the notion of action, but it provides the additional constraint that agency is action that arises out of the animal's nervous system. This seems plausible, but it is not sufficient and perhaps not necessary for agency. Many things arise out of an animal's nervous system: sweating, hiccups, seizures, reflex movements, and so on. Of course, none of these things are proper actions. But this takes us right back to my first point: What is action? On the other hand, the condition (action that arises out of the animal's nervous system) does not seem necessary for mental agency and shared agency. Mental agency comprises things such as making a decision or trying to remember something. It does not seem correct to say that mental acts arise *out of* the nervous system—at least not in the sense in which movements arise out of the nervous system. Shared agency arises when agents act together as a group. This kind of agency does not arise out of any *one* nervous system. Similar worries apply to the third characterization. The whole animal system may undergo various changes that are not agency (sweating, hiccups, and so on). Further, mental agency does not seem to involve the whole animal system, and shared agency does not arise from one particular animal system.

This just shows, I take it, how difficult it is to come up with one definition or characterization of agency that fits all cases. Given this, it is only plausible to distinguish between *kinds* of agency. This is also what Klemm does when he distinguishes the agency of "higher animals" from simpler forms of agency (pp. 2–3 in MS). The distinguishing feature, according to Klemm, is that only higher animals act in accord with *intentions*. This is in line with the philosophical conception of action, to which I turn now.

As mentioned, it is plausible to hold that there are different kinds of agency. One definition of a kind of agency can be derived from the account of action that is widely shared and taken for granted within philosophy. On this standard view, all actions are intentional *under some description* (Anscombe 1957; Davidson 1963). The easiest way to explain this, without going into the technical details, is by way of an example. As I type these words on my computer, I am wearing down the keys of my keyboard. Typing these words is an intentional action of mine. Wearing down the keys is also something that I do, but it is not an intentional action. What makes this (wearing down the keys) an action, according to the standard view, is the fact that it is intentional under another description (namely, that of typing these words). So, on this view, intentionality distinguishes actions from other movements, behaviors, or events (such as slipping, falling, sweating, coughing, and so on). Further, most

versions of the standard view explain intentionality in terms of the agent's mental states and events and in terms of their causal roles. Very roughly, some movement (or event) is intentional, on this view, if it is caused and guided by an *intention,* and if this intention is based on the agent's *reasons* in the minimal or subjective sense that having the intention is caused and rationalized by the agent's desires and beliefs (for more on this, see, for instance, Enç 2003; Mele 2003).[2]

This philosophical account of action yields a straightforward definition of a higher kind of agency: intentional agency. On this approach, the exercise of this higher kind of agency consists simply in the performance of intentional actions (as defined above). This raises the question of what simpler forms of agency consist in, which I shall not discuss here. Let me suggest only that the notion of *goal-directedness* provides perhaps the best starting point here: It seems that an animal's movements can be goal directed even if they are not in any clear sense based on intentions, desires, and beliefs (for more on this, see Barandiaran et al. 2009, for instance).

With this as a background, let me now turn to Klemm's characterization of the process of intentional and conscious agency. On a number of occasions, Klemm points out that intentional and conscious agency is a temporally extended process, and he suggests that this process begins with an intention and ends with the execution of the action. Roughly, he suggests the following sequence (see p. 3 and 29 in MS):

> Intention, evaluation, decision, planning, execution.

Let me first point out here that this is out of line with the philosophical conception of intentional agency. On this conception, intentions are based on the agent's reasons, which are usually construed as the agent's desires and beliefs (or as the things that are represented by the agent's desires and beliefs). This would seem to correspond to the element of evaluation in Klemm's account. And this would mean that the two views disagree on the *order* of the elements in the sequence. Further, Klemm pulls apart intention and decision, whereas most philosophers hold that making a decision just *is* forming an intention. On their view, deciding to go the cinema tonight is nothing over and above forming the intention to go to the cinema tonight, for instance. Of course, once you have decided to go to the cinema, you have to make further decisions on what to see, how to get to the cinema, and so on. You have to make, what we may call, sub-decisions, which consist in the formation of sub-intentions (on how to the implement the goal). These sub-intentions fill out the further details concerning the means or the manner of attaining the goal. Given that such sub-intentions are usually also based on reasons, we can reconstruct the philosophical conception of the process as follows:

> Reasons (desires, beliefs, evaluation), decision (formation of an intention), sub-decisions concerning the means (formation of sub-intentions based on further reasons), execution.

[2] Note that this view applies also to mental and shared actions, provided that we can form some mental states intentionally and provided that the members of a group can be said to have shared intentions, goals, and beliefs (all of which might be based on agreement or some kind of voting system).

It seems that the difference between the two views is, in essence, the following. On Klemm's view, decision making concerns the question of how to implement an intention that stands at the beginning of the processes. According to the philosophical conception, reasons are at the beginning of the process.

I would like to make two points here. First, it is worth noting that most scientists who work on neuroscientific models of economic decision making use a conceptual framework that is closer to the proposed philosophical account of the sequence than to Klemm's. In particular, most models in neuroeconomics take also reasons (values or preferences) as the starting point of the sequence (for overviews and reviews, see Glimcher et al. 2009). Second, even though Klemm is very critical of the well-known (and notorious) neuroscientific experiments on free will (Libet 1985; Soon et al. 2008), his own reconstruction of the process of agency seems to exhibit one of their shortcomings, because these experiments also overlook or neglect the point that intentions and intentional actions are usually *based on reasons*. (I will say more about this in the following section.)

7.2 The Neuroscientific Study of Free Will

Most psychologists and neuroscientists seem to think that the belief in conscious agency and free will is illusory. Klemm is highly critical of their claims (see propositions 7 and 8, pp. 16–23 in MS), and I agree with most of what he has to say about this (for more on my take on these issues, see Schlosser 2012a, b, 2013, 2014). But, I also think that Klemm does not go quite far enough in his critique, and I would like to raise some more general conceptual and methodological issues here.

I agree with Klemm (p. 19 in MS) that the choices that are studied in the Libet experiment (1985) and in the follow-up experiment by Soon et al. (2008) are not representative of the choices that we make in our everyday lives—they are, at least, not representative of the more significant choices for which we hold each other responsible. However, when Klemm goes on to offer some further comments on this, he assumes a model of decision making in which the expected utilities of competing courses of action are evaluated before the final decision is made (see, in particular, the caption to Fig. 4.3, p. 21 in MS). It seems to me, however, that one problem with the mentioned experiments is that such a model of decision making does not apply at all, because participants are not presented with any real alternatives that can be evaluated or ranked. In the Libet experiment, participants are asked to perform a predefined movement when they feel like doing so. In the Soon et al. experiment, they are asked to press a button with either their left or right index finger when they feel like doing so. In both cases, participants have absolutely no reason to move now (rather than at some other time) and no reason to use one index finger (rather than the other one). Of course, in a sense they do have alternatives. But they do not have real alternatives in the sense that the options are indistinguishable in terms of their value and in terms of their consequences. There is simply nothing that can be compared and evaluated,

and so participants *cannot even begin* to engage in a process of *proper* decision making. So, not only are those choices not representative. They are not based on processes of proper decision making at all.

This assessment is fully in line with the model and the experiment provided by Schurger et al. (2012). According to their model, the decision in the Libet experiment is not based on any evidence (value or reason) at all. It is, rather, determined by random fluctuations in neuronal activity. They tested this model by means of a simulation of the Libet experiment and they conducted an experiment which confirmed the model. As they point out, their model is also consistent with the existence of the kind of pre-decision biases that were found in the Soon et al. experiment, which may "reflect stochastic fluctuations rather than an intentional (pre-conscious) decision-process" (Schurger et al. 2012, p. 6). Moreover, there is simply no reason to think that this stochastic model of decision making applies to other tasks and ordinary decisions, because in other tasks and in ordinary situations there is usually some evidence (value or reason) that the agent takes into account.

This brings me to a second point about the neuroscientific experiments on free will (and voluntary action). There is a very widespread trend or tradition in neuroscience to define free will (and voluntary action) by contrasting it with actions that are triggered or driven by external causes. I do not know whether Libet was the first who operationalized free will in this way. But this dichotomy between free (and voluntary) versus externally triggered (or driven) action plays a central role in Libet's argument against free will, and it provides the operational definition that underlies the design of the Libet experiment and of numerous other experiments on voluntary action thereafter (see Libet 1985; Jahanshahi and Frith 1998; Dreiber et al. 1999; Jenkins et al. 2000; Haggard 2008; Passingham et al. 2010; Hughes et al. 2011, for instance). An explicit commitment to this approach can be found in the review article by Patrick Haggard, who writes that a "scientifically […] satisfactory approach defines voluntary action by contrasting it with stimulus-driven actions" (2008, p. 934).

Elsewhere, I have argued at length that this conceptualization of free will is deeply flawed (Schlosser 2014). I will not try to summarize the full argument here, as my main point can be made effectively by means of an example. Suppose that you are sitting at your desk, working on something. At some point, the phone rings. Depending on the particular circumstances, and depending on your habits, you might respond in different ways. You might, for instance, immediately pick up the phone, perhaps because you have the habit of doing so. This does not mean that you would always respond in this way. For instance, if you have an urgent deadline to meet, you might either ignore the phone or you might pause for a moment and briefly consider whether or not you have the time to talk to someone right now. Comparisons between such possibilities support some important observations. First, in some cases, you may respond habitually or automatically, whereas in others, you may respond after a brief moment of deliberation. But, in each case in which you pick up the phone, you respond to

an external factor (cause, cue, or trigger). This highlights a first shortcoming of the mentioned dichotomy between free (or voluntary) versus externally triggered (or driven) actions. There are significant differences between different ways of responding to external factors: They are all *responses* to something, but they are not all *triggered* in an automatic fashion. Moreover, even if a response is triggered, it is simply not obvious that it is therefore involuntary or unfree. Suppose, for instance, that you are expecting an important phone call. Because of this, you might immediately pick up the phone as soon as it rings, and you might have the intention to pick up the phone as soon as it rings. In a sense, at least, your response would be triggered by an external cue. Does this mean that picking up the phone would therefore be involuntary or unfree? I do not think so. It is, at least, far from obvious that this response would be involuntary or unfree, especially if we take into account the fact that it might be *based on a prior intention* to respond quickly.

Given this, any approach that defines free will (or voluntary action) by contrasting it with actions that are performed in response to external factors would appear to be flawed. To take another example, suppose you take an umbrella in the morning in response to seeing dark clouds over the sky. This appears to be a free and voluntary action, and it may well be a free and voluntary action even if you did not consciously deliberate about whether or not to take an umbrella. It is, I contend, simply a mistake to assume that free choices (and voluntary actions) must not have external causes. In fact, some reflection on everyday decisions suggests that our choices and actions should usually have external causes, because they should be responsive or sensitive to external factors (such as dark clouds over the sky). Choices and actions that are altogether insensitive to environmental circumstances do even seem dysfunctional and random. Given this, it is rather unfortunate that the existing neuroscience of free will (and voluntary action) is largely about such choices—choices that are not based on any reasons and that are not made on the basis of anything that has significance or value.

It should be clear that this is not merely a conceptual or semantic issue. Far from it, the operational definition of free will and the design of the neuroscientific experiments are based on this problematic conception of free will. To his credit, Klemm does not reproduce this mistake. He says, for instance, that

> […] willed action requires some input to the brain circuits that generate willed action. Such input may come from an external contingency or may be generated internally from some emotional drive or motivation. (p. 20 in MS)

This suggests that, on Klemm's view, free and voluntary actions may well have external causes, and so it seems that Klemm departs from the common neuroscientific practice of defining free and voluntary actions by contrasting them with externally caused actions. This point deserves emphasis, because Klemm himself does not make this explicit, and because this point really does, in my opinion, uncover a very serious conceptual and methodological shortcoming of the neuroscientific experiments on free will and voluntary action.

7.3 Metaphysical Presuppositions: Dualism and Incompatibilism

In the two sections on the scientific challenges to conscious will and free will (pp. 16–23 in MS), Klemm addresses some issues that are connected to the mind–body problem and to the question of whether free will is compatible with determinism, which are traditional philosophical issues. I largely agree with the points he makes here, but I would nevertheless like to add a few remarks and observations.

When psychologists and neuroscientists draw their radical conclusions about the illusion of conscious will and free will, they sometimes presuppose dualism about the mind–body problem (Libet 2001; Haggard & Libet 2001; Wegner 2002) and they often presuppose incompatibilism about free will and determinism (Libet 2001; Haggard & Libet 2001; Haynes 2011). Again, Klemm does not seem to share these presuppositions, and again he deserves credit for this.

What if, Klemm asks (Chap. 4, in this volume, p. 16 in MS), consciousness is constituted by neural events? If consciousness is constituted by neural events, then it does not have to intervene, somehow, in neural processes. In my opinion, that is exactly the right starting point for thinking about the role of consciousness in the initiation and guidance of action; if consciousness is to play a role, we better construe it as something that is constituted or realized by neural events (states or processes).

Further, Klemm makes some remarks, which suggest that he is sympathetic to compatibilism about free will and determinism. It is not clear to me whether he meant to suggest this, and I do not know whether or not he is a compatibilist. But when he asks, rhetorically, whether we are slaves of reason when we make wise choices (p. 23), he certainly sounds like a compatibilist to me. Compatibilists are always keen to point out that not all kinds of causal determination rule out free will. They argue, in particular, that our choices may well be free if they are determined by our *reasons,* because determination by reasons is *persuasion,* not *coercion*. Klemm further notes that we would lack free will if our choices were immutable (p. 22). Again, this is something that compatibilists like to stress, because an agent (organism or system) that is causally determined need not be immutable at all. Determinism is perfectly compatible with development, learning, and frequent change.

No matter whether or not Klemm is a compatibilist, it is worth noting that some of his arguments and remarks are fully in line with compatibilism. More importantly, it is worth stressing that other neuroscientists should consider compatibilism more carefully as well. They should, at least, be aware of the fact that their conclusions about free will are often based on an unquestioned and unjustified assumption of incompatibilism just as much as they are based on empirical evidence.

7.4 The Nature and Role of Consciousness

As mentioned, Klemm suggests that consciousness is constituted by neural events. He holds, in particular, that it is constituted by certain circuit impulse patterns (CIPs). This leaves open the question of how we should think about consciousness to begin with. As Klemm notes, there seem to be two basic options (p. 24 in MS): We can think of consciousness as a certain type of *state*, or we can think of it as a certain type of *being*. Klemm does not consider the first option. He advances, instead, a version of the second: "consciousness is a nerve-impulse based brain state existing as a being that acts in the world" (proposition 10, pp. 24–26 in MS). I must say that I found it rather difficult to follow Klemm on this—both in the sense that I found it difficult to understand the view fully and in the sense that I found myself in disagreement with most of the things that I did understand. Before we turn to that, let me offer a brief remark on Klemm's representation of Descartes' view on consciousness.

According to Klemm, Descartes thought that the contents of consciousness are presented "on the stage of a 'Cartesian Theatre' for viewing by a virtual little man," and he goes on to say that this idea is nowadays "ridiculed by most scholars" (p. 24 in MS). Descartes, I should like to note, never held such a view and he never wrote anything that implies it. In fact, Klemm himself inadvertently ridicules Descartes' view by misrepresenting it in this way. Descartes never said or implied that consciousness is a "little man" and he never used the metaphor of a stage on which the contents of consciousness are presented for viewing. Yes, some philosophers (and scientists) talk that way about Descartes' views *in order to* ridicule them. But, this should not be taken to mean that he actually said such things.

At the core of Klemm's own view is the notion of an *avatar*, which is meant to replace and, in a sense, rehabilitate the notion of a "virtual little person" (or homunculus). An avatar is construed as a "being with agency": avatars "sense, evaluate, decide, and initiate and direct action" (p. 25 in MS). It is suggested that we should think about "consciousness as an avatar, generated by the brain as a set of CIPs to act on behalf of the best interests of the brain and body" (p. 25 in MS).

Let us first ask how this notion of an avatar can be combined with Klemm's naturalistic commitments—in particular, the commitment to the view that consciousness is constituted by neural events and processes (CIPs). On this, Klemm says that an avatar

> […] is a CIP representation of the body and what goes on inside and outside of the body, all referenced to the sense of self, which itself is a CIP representation of the "little person". I contend that the brain creates a conscious homunculus in the form of an avatar that it deploys to act on behalf of the embodied brain in ways not otherwise possible. (Klemm Chap. 4, in this volume, p. 25 in MS)

I find this rather difficult to understand. In particular, it seems to me that the notion of representation creates confusion here. Why is the sense of self a representation of a "little person"? If by sense of self we mean bodily awareness, then we can understand the sense of self in terms of perception, proprioception, and body schemata. If by sense of self we mean having an individual and narrative identity, then we can

understand it in terms of the contents of cognition and metacognition (beliefs about who I am, where I come from, and awareness of having those beliefs). If, finally, by sense of self we mean a sense of agency, then we can understand it in terms of having conscious intentions and in terms of anticipated and perceived feedback from the consequences of our movements. At no point, it seems to me, are we aware here of a "little person" when we have a sense of self or a sense of agency.

Further, it has not become clear to me what Klemm means by the claim that the avatar is acting *on behalf* of the embodied brain. According to the view, an avatar is consciousness construed as a conscious being with agency. But why should I think of my consciousness as acting on behalf of my embodied brain? This strikes me simply as a very odd thing to say. In my opinion, the problem with this suggestion stems, ultimately, from Klemm's decision to view consciousness as a being (entity or agent). If we take this as our starting point, then it is difficult to avoid the implication that I am actually *two* things: an embodied brain and a conscious agent. If I am, in this sense, two things, then it does make sense to say that one of the two is acting on behalf of the other. But for all I know, I am not two things, and my consciousness is not acting on behalf of my embodied brain when I am acting.

A much more plausible starting point is the assumption that conscious agency consists in the fact that conscious states or processes of mine, such as having or forming conscious intentions, play certain causal roles in the initiation and guidance of my actions. On this view, consciousness is not a being (entity or agent), but it consists in having conscious mental states that play certain causal roles. It is widely agreed that many mental states or processes are unconscious. When consciousness comes into being, it is not the case that a new being or agent comes into existence. Rather, when consciousness comes into being, certain mental states or processes become conscious. And I engage in conscious agency when such states or processes initiate and guide my actions. Of course, this view raises many questions and problems of its own, which are beyond the scope of this commentary. Let me stress, however, that it does avoid at least two of the issues that arise for the view proposed by Klemm. First, as the sketched view does not stipulate another being (entity or agent) apart from the embodied brain, it avoids the odd suggestion that consciousness is acting on behalf of the embodied brain. Second, it construes conscious agency in terms of the causal roles of conscious mental states. It offers, thereby, an explanation of what conscious *agency* consists in. In contrast, Klemm's view does not really explain conscious agency, because it refers to an entity, an avatar, which is *defined* as a conscious being *with agency*. This view, in other words, presupposes the notion it seeks to explain.

7.5 Concluding Remarks

In reply to critics, Benjamin Libet once made the following two points. First, he claimed that he had fully considered the implications of his experimental findings for the concept of free will (Libet 2002, p. 292). Second, he noted that the negative

criticisms of the findings and their implications have come mostly from "philosophers and others with no significant experience in experimental neuroscience" (p. 292). Libet, I think, committed two mistakes there. He did not fully explore the implications of the findings, as his conclusions about free will were based on an idiosyncratic and rather problematic definition of free will and on an unquestioned commitment to dualism and incompatibilism. Further, as I have pointed out, experimental designs are based, in part, on operational definitions, and operational definitions are based on conceptual frameworks. Yes, philosophers have usually no experience in experimental neuroscience, but they have experience in assessing the coherence and plausibility of conceptual frameworks and definitions. So, contrary to what Libet seemed to imply, the comments and criticisms from philosophers can be relevant to neuroscience, insofar as they can inform operational definitions. Klemm's chapter is to be commended for avoiding such mistakes. He does not define free will and voluntary action simply by contrasting it with externally driven action, and he explores a way of thinking about free will that departs from the unexamined commitment of other neuroscientists to incompatibilism.

To conclude, let me point out what I take to be the two most important points for future research on human agency and free will (from a philosophical point of view). First, I think scientists should reconsider the operational definition of free will (and voluntary action) in terms of the dichotomy between free (or voluntary) versus externally driven action. Second, before drawing radical conclusions about conscious agency and free will, scientists should consider the fact that such conclusions are usually based as much on metaphysical presuppositions (such as dualism and incompatibilism) as they are based on experimental results. As pointed out, Klemm avoids the corresponding mistakes, but he does not, in my opinion, emphasize the underlying issues strongly enough.

References

Anscombe, E. (1957). *Intention*. Oxford: Basil Blackwell.

Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior, 17,* 367–386.

Brass, M., Lynn, M. T., Demanet, J., & Rigoni, D. (2013). Imaging volition: What the brain can tell us about the will. *Experimental Brain Research, 229,* 301–312.

Cisek, P., & Kalaska, J. F. (2010). Interacting with a world full of action choices. *Annual Review of Neuroscience, 33,* 269–298.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60,* 685–700.

Deiber, M. P., Honda, M., Ibanez, V., Sadato, N., & Hallett, M. (1999). Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: Effect of movement type and rate. *Journal of Neurophysiology, 81,* 3065–3077.

Desmurget, M. (2013). Searching for the neural correlates of conscious intention. *Journal of Cognitive Neuroscience, 25,* 830–833.

Desmurget, M., & Sirigu, A. (2009). A parietalpremotor network for movement intention and motor awareness. *Trends in Cognitive Sciences, 13,* 411–419.

Enç, B. (2003). *How we act: Causes, reasons, and intentions*. Oxford: Oxford University Press.

Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E., & Culham, J. C. (2011). Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *Journal of Neuroscience, 31,* 9599–9610.

Glimcher, P. W., Fehr, E., Camerer, C., & Poldrack, R. A. (Eds.). (2009). *Neuroeconomics: Decision making and the brain*. London: Elsevier Academic Press.

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience, 9,* 934–946.

Haggard, P., & Libet, B. (2001). Conscious intention and brain activity. *Journal of Consciousness Studies, 8,* 47–63.

Haynes, J. D. (2011). Beyond Libet: Long-term prediction of free choices from neuroimaging signals. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 85–96). New York: Oxford University Press.

Hughes, G., Schütz-Bosbach, S., & Waszak, F. (2011). One action system or two? Evidence for common central preparatory mechanisms in voluntary and stimulus-driven actions. *Journal of Neuroscience, 31,* 16692–16699.

Jahanshahi, M., & Frith, C. D. (1998). Willed action and its impairments. *Cognitive Neuropsychology, 15,* 483–533.

Jenkins, I. H., Jahanshahi, M., Jueptner, M., Passingham, R. E., & Brooks, D. J. (2000). Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. *Brain: A Journal of Neurology, 123,* 1216–1228.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529–566.

Libet, B. (2001). Consciousness, free action and the brain. *Journal of Consciousness Studies, 8,* 59–65.

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition, 11,* 291–299.

Mele, A. R. (2003). *Motivation and agency*. Oxford: Oxford University Press.

Momennejad, I., & Haynes, J. D. (2012). Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. *NeuroImage, 61,* 139–148.

Passingham, R. E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: From self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences, 14,* 16–21.

Schlosser, M. E. (2012a). Causally efficacious intentions and the sense of agency: In defense of real mental causation. *Journal of Theoretical and Philosophical Psychology, 32,* 135–160.

Schlosser, M. E. (2012b). Free will and the unconscious precursors of choice. *Philosophical Psychology, 25,* 365–384.

Schlosser, M. E. (2013). Conscious will, reason-responsiveness, and moral responsibility. *Journal of Ethics, 17,* 205–232.

Schlosser, M. E. (2014). The neuroscientific study of free will: A diagnosis of the controversy. *Synthese, 191,* 245–262.

Schurger, A., Sitta, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences, 109*(42), E2904–E2913.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*, 543–545.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge: MIT Press.

Chapter 8
Starting Points for Agency Research

William R. Klemm

I am pleased with the generally supportive response from all three commentators (Zhu, Toates, and Schlosser; Chap. 6, this volume). It was particularly gratifying to see so few challenges to the list of axioms and propositions. Perhaps these can now serve as a starting point for research and scholarly analysis of agency. Each commentator made important comments that clarify and expand on the issues. I respond to each as follows:

8.1 Agency in Life (Zhu; Chap. 6, this volume)

For all animals, agency is the key to successful life, and this commentator makes it explicit by saying we should put "agency in a creature's ecological niche and whole life realm." This commentator and another rightfully focused on what I said were the elements of human agency:

> Intend, remember, value, decide, prepare/plan, and act.

This author said that these elements are correct and promising, and emphasized that not only do we need to specify elements of agency but we should also understand structures of agency, namely, how elements of agency are related to each other and organized. That is exactly why I took pains to identify them. One can quibble about serial order, as the next commentator does, but we have to start somewhere, and this seems to be the place to start.

This author emphasizes, as I would agree, the reasoned planning component. This seems particularly important for conscious human agency, which is particularly adaptive because planning can be evaluated explicitly in consciousness. This goes to the heart of the argument of many that consciousness does not do anything.

W. R. Klemm (✉)
Texas A&M University, College Station, USA
e-mail: wklemm@cvm.tamu.edu

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_8

Certainly, one's brain can plan actions unconsciously, but consciousness provides a robustness of logic and analysis, not otherwise achievable.

This author recognizes the flaws in Libet-paradigm experiments on free will. Indeed, the other two commentators are not enamored with this research either. One of my major contentions is that the brain is doing a whole lot of things besides deciding when to press a button. I have referred to these multiple ancillary thoughts as thinking about the "rules of the game," and this reviewer elaborated the point with some salient new examples.

The author's point about the fallacy called "consciousness localization" is quite interesting. All mental states and processes are divided into two categories: conscious and unconscious. The nature of the fallacy is the neglect that consciousness is in the first place a feature of the whole agent. Thus to say that a given behavior results from a conscious or an unconscious state probably is a fallacy. Neither state is independent of the other. Both may be involved in varying degrees in every act of agency in those species capable of consciousness. See my just-released book, *Mental Biology: The New Science of How the Brain and Mind Relate*.

8.2 Agency and Free Will Commentary (Schlosser; Chap. 6, this volume)

Notably, this author self-defines as a philosopher and concludes that from a philosophical perspective that I must be a compatibilist. In terms of free will, I take the stance that the neuroscience on this matter is flawed, and perhaps there is no satisfactory way to resolve this issue through experimentation. If so, philosophy and religion are all that we have left. Philosophy does have much to say about the importance of operational definitions, and this author and I both agree that neuroscience has approached free will issues with rather sloppy definitions.

I thought it particularly interesting that the matters of executive control are seldom part of philosophy, whereas in recent years it is becoming prominent in neuroscience. Also interesting is this author's rejection of the elements of agency that I identified as intention, remember, evaluation, decision, planning, and execution. Perhaps the mind-set of philosophy accounts for this difference. One point I would make is that intention and decision are separable, not as this author claims, that decisions are just forming an intention. One can have an intention to eat, for example, while still needing to decide what to eat. Of course, once I decide to eat blueberry pie, then the intention and decision do become conflated.

As with the "Agency in Life" commentator, this author supports my arguments about flaws in free will research and provides even more reasoned argument. I find, however, problems with the claim that in Libet-type experiments my model of decision making is not relevant. It is stated that my model's sequence (intend, remember, value, decide, plan, and then monitor what happens) does not apply because the subject has no choice. Yes, he does have choices. He has to remember and follow

the rules of this experimental game. He has to intend to make a button press and pay attention to the clock and to what he thinks is the instant of conscious realization. The point this commentator makes is that these are not real alternatives, because there is no value element. True, but that is not the whole truth. The subject still makes a value judgment: "Is now a good time to press? How long has it been since the last press? Am I dragging this thing out too long? Am I rushing? Does it matter how long I wait? Will I mess up the experiment if I am sloppy about noting when I actually made the decision or where exactly the clock hands were?"

Perhaps both of us need to reevaluate this sequence of agency elements. They may not always appear in simple serial order. You may, for example, think about a trip to the grocery store (plan), prefer (decide) to buy blueberry pie, and then realize (remember) that you can buy another kind of pie for less (value). Perhaps the commentator and I have gotten "into the weeds" as they say. Philosophy can do that.

I am particularly gratified that this commentator agrees that the right starting point for the role of consciousness is the fact that consciousness is constituted by neural events and does not have to intervene in neural processes because consciousness is a neural process.

This commentator takes issue with my proposition 10, that consciousness is a state of being. His problem with my avatar idea seems to be my claim that the avatar is a neural representation of a "little person." That idea of "little person," in terms of physiology is the topographical map, both sensory and motor, of the body. The brain knows its body, where everything is, and how to engage its parts in agency. To the brain, there is a little person (body) inside. I disagree with the assertion that we are not aware of a "little person" when we are self-aware or initiating agency. Of course, the brain thinks of itself in terms of its body.

Why should we think of the avatar of acting on behalf of the embodied brain? I think that it is this conscious avatar that gives us such adaptive power as a species. Conscious mind can enrich and perfect the drives and fuzzy thinking of unconsciousness. This commentator hits on something important when interpreting my position as advocating a split brain, of two beings in one brain. I regard unconscious mind and conscious mind as an interacting unity, not as the commentator thought that I claimed consciousness to be apart from embodied brain. Unconscious and conscious processes inform each other, but in different ways. Let me also point out that once one accepts that consciousness is a state of being, it is easier to defend the position we both have that humans have a degree of free will (executed by the conscious state of being).

8.3 System I and II Classification (Toates; Chap. 6, this volume)

This commentator, I think, provides a helpful way to think about agency in terms of its goal-directed nature and its susceptibility to reinforcement. These features of behavior have been studied for a long time, but not so much in the context of

agency. All of us seem to agree that agency provides a useful umbrella way to think about behavior. Indeed, this book might have been conceived by the editors with that in mind.

The author alludes to the historical class of worldviews of operant conditioning and cognitive neuroscience. Neither of us apparently wants to waste time rehashing those views and agree that these world views need to be integrated into the concept of agency.

Defining agency is not resolved. One can argue, as this author does, that command neurons in invertebrates, for example, should not be considered sources of agency. Certainly, human intention, memory, value, etc., are expressions of agency at a much higher level. Does agency only occur in the conscious mind? Well, it all depends on how you want to define it. I defined agency more inclusively because all animals must act in the world to survive. Such action is prerequisite for animal survival. Animals are not trees.

This commentator refers to the model of two systems: system I (fast, reactive agency, and unconscious) and system II (slower, reflective, and operating in consciousness). Though not explicitly stated, this distinction lies at the heart of objections to Libet-type free-will experiments, which entailed mostly system I processes, and should not therefore be interpreted to explain system II functions. The author asks whether I associated agency with systems I and II, rather than just system I. Yes, I do.

The commentator makes a useful analogy about neural correlates, wherein it is stated: “Does the flow of sodium and potassium ions cause the action potential? According to identity theory, presumably it does not, for there is not a time sequence of, first, influx of sodium and then the action potential. Rather the action potential is the flow of ions,….” This dovetails exactly with my position about the circuit-impulse-pattern (CIP) basis of conscious being. A certain set of CIPs are not so much causes of consciousness as they are consciousness itself.

The comments about “top-down” and “bottom-up” remind me of why I told the editor I did not want to say much in this regard because these terms, though frequently used, are ambiguous and not particularly helpful.

This commentator was brave in talking about “ghost in the machine.” No serious scholar denies that conscious mind resides in the brain. Does it reside only in the brain? Science cannot test that. Indeed, there is a lot of “spooky physics” (general relativity, quantum mechanics, string theory, parallel universes, dark matter, dark energy) that raise the possibility that what we traditionally regard as spiritual might just be material realities not yet discovered. It may be time for neuroscience to consider a paradigm shift, at least to a point of not dismissing noetic science out of hand. As a result, I have created a college course at Texas A&M University in Neuroscience and Religion. Student interest is very high.

Chapter 9
Neurobiology of Agency: "Conatograms" and the Ghost in the Machine?

Matthew G. Clark

This section explored and challenged existing conceptions of agency in humans and other animals. More directly, we endeavored to initiate direct examination of the theoretical relevance of agency specifically with regard to psychobiology and philosophy. The objective was to present an accessible, axiomatic, and functional conception of the neurobiology of agency in everyday life in humans. From the outset of this volume, Roger Smith clearly presented that the operational definition of agency is central to its evaluation and interpretation in science. He further demonstrated that it is a poorly defined and decidedly un-unified concept across the scientific community, and particularly across psychology. Therefore, he proposed a conception that defines agency in modern times as "causal processes (even causal determination) and free action." While this definition covers a great deal of behavior, his intent appears to have been to start to shape the dialog and define the historical context in which agency has existed until now. This is clearly also an issue for those studying the neurobiology of agency as it was raised by all contributors to this section. As you will see below, these contributors would likely support Smith's definition related to "causal processes of action," as long as the term "free" was removed. This is not necessarily because the commentators do not believe in "free will," they just do not seem to agree that this term is scientifically useful at this time.

Neuroscience has grown immensely in popularity over recent decades and scientists from this subdiscipline have successfully described and ascribed neural activity to various constellations of human behavior. We assembled a collection of neuroscientists and philosophers in order to create axioms of agency and potentially build bridge laws between the eliminativists (and related scientists), "folk" psychologists, philosophers, and others, and possibly even address both compatibilist and

DISCLAIMER: The views presented in this manuscript are those of the authors and do not purport to reflect the positions of the funding agencies, the United States Military Academy, the Department of the Army, the Department of Defense, or the U.S. Government.

M. G. Clark (✉)
United States Military Academy, West Point, NY, USA
e-mail: Matthew.Clark2@usma.edu

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_9

incompatibilist perspectives directly through the prudent application of published empirical data and reasoned thought. While this synthesis will not reexamine all of the contributors' perspectives, we will explore some of the more salient points and where they interact. This synthesis also introduces a few more thoughts on the neurobiology of agency for your consideration. Including the neurobiology of agency in this volume was an essential requirement for full consideration of this topic. Fortunately, the "tent" of psychology is sufficiently large to house all natural sciences alongside other sociocultural, philosophical, or political perspectives. Exploring a concept such as agency requires broad consideration.

By academic tradition, neuroscience seeks to reduce humans and other animals to the biological elements that guide and direct all actions and behaviors. The result is that behaviors and the neural correlates are often overly reduced to the point where complex coordinated behavior is not evident in the individual action potentials and collective patterns that created the action. Yet, neuroscientists still seek relevant neural correlates or representations of all aspects of the human condition. This is akin to studying atmospheric chemistry and physics with the goal of predicting the weather. These scientists work to analyze complex phenomena through their constituent parts to eventually understand why, how, and when it will rain, snow, or even when there might be a tornado or hurricane. This is admittedly oversimplified, but this is exactly the same relationship neuroscience has in concept with the broader field of psychology. A similar example in psychology is the active search for the elusive "engram," which was started by Richard Semon and Karl Lashley in the 1920s (Josselyn 2010; Lashley 1950). In concept, the hope was that they might find an engram in neural activity that characterized the neural correlates or representation(s) of memory.

Likewise, neuroscience now seeks an even more elusive "conatogram" within the recesses of the brain as the neural correlates or representation(s) of agency. I have chosen to call this a conatogram because conatus and conation have had a storied historical past in scientific and philosophical thought. Thomas Hobbs captured the nature of this quest in his desire to simplify and explain cognition and the mind through "conative functions" (Bidney 1962). Therefore, we can see that modern neuroscience seeks a conatogram or conatograms in essentially the same fashion as Lashley, only the focus is on human agency and will. There is a belief among neuroscientists that such a conatogram must exist, otherwise, why keep searching? Put another way, agency, consciousness, and potentially will all arise out of the brain. If this is accepted, then there must be some type of neural representation. While equally elusive, the expectation is that such a conatogram will either expose or uncover the presence or absence of a "ghost in the machine."

9.1 Ghosts, Zombies, and Libet's Experiment: The Illusion of Agency

Most often, neuroscience is broadly painted as suggesting that there is simply no ghost in the machine, or perhaps that the machine is a zombie (Koch and Crick 2001). Replicated several times, this conclusion is often attributed to a study by

Benjamin Libet and colleagues that concluded voluntary motor behavior can be attributed to an unconscious "readiness potential" (Libet et al. 1983). This experiment was not directly investigating agency or free will as is often overstated; it was instead measuring voluntary motor behavior and the conscious intent to perform a simple movement of the hand. Given that each of the contributors to this section of Volume 12 of *Annals* mentioned this classic study, it is useful to briefly describe the experiment.

In Libet's experiment (Libet et al. 1983), subjects were asked to relax in a lounge chair and initiate the study when they were "comfortably ready." Thereafter, each was presented with a tone signaling 1–3 s of preparation time where each fixed his or her gaze on a rotating oscilloscopic timer. With their gaze focused on the center of the timer, at the end of the preparatory period, the 2.56 s timer started a clockwise rotation. At a time of the participant's choosing, he or she was to make a quick movement of their hand (i.e. flexion of fingers and/or wrist of the right hand). The participants were asked to report the time that they felt the spontaneous urge to move while simultaneously electroencephalographic (EEG) activity was recorded in the frontal lobe (recorded at the vertex) and synchronized to this "type II" (i.e. reasoned to be spontaneously generated) behavioral task. At least at the level of the frontal cortex (perhaps specifically around the presupplementary motor area), the results showed a "readiness potential" that preceded the self-reported urge or intent to move by approximately a third of a second and the actual movement occurred another 200 ms later.

As Smith implied earlier in this volume, the standard belief within neuroscience is that agency is largely determined and that the sense of agency is illusory. Libet's experiment is some of the main evidence supporting this conclusion because of the underlying belief that conscious awareness is a requirement for voluntary behavior, a sense of agency, and ultimately free (nor not so free) will. This position is further supported by other research showing neural activity preceding conscious awareness by up to 10 s prior to the actual motor behavior (cf. Soon et al. 2008), which is an exceptionally long period of time relative to brain activity. The natural conclusion is that because conscious awareness follows in time, consciousness must not play a role—it is only an observer and reporter.

9.2 Agency as Circuit Impulse Patterns

Klemm, our first contributor, seeks to break the concept of agency down into its constituent parts by presenting ten axioms and propositions in a way that addresses this topic more completely. His approach is extremely useful given the complexity of the topic. He specifically defines agency as the ability of an animal "to act in the world" noting that "the behavioral expression of agency results from [neural] mechanisms that enable or promote it, as well as those that suppress it." He uses comparative biology to address agency in various animals with different levels of neural complexity to demonstrate the potential activity and structures involved in the following cognitions or behaviors that he believes comprise agency:

Intention, Remembering, Valuing, Deciding, Prepare/Planning, and Acting

Inherent in his description of the neural representation of agency, Klemm argues that we should start our search for relevant activities by examining neural oscillations, feedback loops, and reentrant mapping. He contends that as higher order animals, we have the capacity to create intentions, values, and plans that ultimately impact neural organization and physiology in individuals. In particular, he contends that agency begins early in our development and that the necessary multilevel circuitry is heavily influenced by our ability to recognize and respond as a self-referential system. To this point, Klemm argues that agency exists within circuit impulse patterns (or CIPs) and that these CIPs are the conatograms scientists seek relative to agency. The problem is that at this point, there is no clear locus that "controls" agency. However, he suggests that because we have not found these CIPs does not mean that they are not there.

Essentially, the CIPs that create agency are diffuse and cannot necessarily be localized to any specific neural structure as psychologists have been so often inclined to do since the days of phrenology and before. While certain elements of agency may be derived from focused neural domains (e.g. executive control is likely involved in consciously directed agency), Klemm suggests that agency is most likely derived from diffuse spatiotemporal patters of CIPs involving various brain regions. In effect, CIPs are thoughts, but they are thoughts that are modified by feedback loops, neural oscillations, and by the structure of the nervous system (presumably within the association cortices and related areas of the brain).

This proposal that agency emerges from diffusely arrayed CIPs across species may be supported by recent research on the development of functional neural networks. Research by Fair et al. (2009) demonstrated that as we develop to adulthood, functional brain networks move from a more local to a more distributed and distal organization. Likewise, in another study, it was reported that as the brain matures, connections between distant centers in the brain become more integrated (Dosenbach et al. 2010). Basically, as humans mature, their brains become more decentralized away from primary behavioral processing centers (e.g. primary motor, visual, somatosensory, emotional cortices, etc.). We also know that executive function, a significant element of agency, continues to develop along a similar time course (Diamond 2013). Presumably, the two are related and it can be predicted that the search for a conatogram for agency will be challenging due to the diffuse nature of the supporting circuitry. Regardless, this evidence supports the perspectives of Klemm and Toates in this volume, and the referenced research specifically supports Zhu's suggestion that neuroscientists avoid the fallacy of "consciousness localization." The published data and these thoughts provide viable targets for further exploration of CIPs as the conatogram for agency.

Essentially, Klemm uses these circuit patterns to suggest that various deterministic processes ultimately provide individuals with agency. It is important to note that this is not necessarily saying agency is a deterministic process, but it is saying that the neurons and possibly even the circuits in the nervous system that produce these patterns are deterministic in their individual behavior. Klemm states that this biological machinery completes the computations in multiple nonlinear processes or networks that then generate agency, perhaps through population-based processes across regions of the brain.

If a main network of agency exists, however, it will not be found in the activity of a simple neural circuit. Instead, the network that accounts for agency is potentially a hierarchical neural network that is centered on mechanisms for executive control and working memory. This means that the best targets for locating this network would be to begin in the dorsolateral prefrontal, parietal, and cingulate cortices. This also implies that different behaviors may have somewhat different executive agents in the brain. For example, it can be said that the motor cortex is the executive agent of motor behavior and the amygdala is the executive agent of emotional behavior. However, both could be subject to processing in a different stage in multiple locations in the brain depending on the given behavior. Importantly, Klemm suggests that intention, planning, and decision making for different behaviors occur in various brain regions for a consciously willed or goal-directed behavior. This is a view that should be the main focus of agency as argued by both Toates and Schlosser (this volume). Therefore, this means that there could be different systems that produce agency depending on the requirements of the behavior and the experiences of the animal, which is further supported in the commentary provided by Toates.

9.3 Systems of Systems that Motivate Agency

Toates essentially argues for differences in neural correlates that focus on novelty and well-rehearsed behaviors by directing much of his commentary to the importance of motives and reinforcement in behavior. He specifically challenges and supports Klemm's conception by exploring a system-of-systems approach in the search for the conatogram. He further suggests that incentive motivation is central to human agency. His commentary would suggest that a fruitful approach for defining and researching agency would be to keep the focus on goal-directed behavior, the relative importance of reinforcement on all aspects of animal behavior (including neural circuitry), and associated elements of agency like incentive motivation. More specifically, he proposes an in-depth consideration of a "System 1 and System 2" classification to avoid potential differences in comparing animals and humans.

As explained by Toates, System 1 is a very responsive system that matures faster than System 2. Evolutionarily, this suggests System 1 is similar in many animals (at least vertebrates) and it largely performs its function in an automatic or unconscious way. This also suggests that the related neural circuitry is evolutionarily conserved and the networks that comprise the system remain in "higher animals." These systems did not change dramatically through evolution because they are central for life. Simple examples might include the neural circuitry for basic behaviors like sleep, attention, and basic life functions within the hindbrain and midbrain. In humans, these behaviors have forebrain involvement as well, but they were conserved as we evolved. For example, in recent years, we have learned that some of the neural structures of the hindbrain also play a role in basic cognition, memory, and other higher order behaviors. An example is the cerebellum, which was long considered to have a role mainly for coordinated movement, but more recent evidence suggests

that the cerebellum plays a role in cognition, memory, language, and even emotion (Baillieux et al. 2010; D'Angelo and Casali 2012; Schmahmann 2010; Thompson 2005). This all supports the idea that the circuits for agency could be diffusely organized.

Similarly, considering Toates' concepts with Klemm's axioms suggests that System 2 developed evolutionarily on top of and tightly integrated to System 1, but it does not supplant that system. System 2 includes the behaviors that are the most associated with the complexity typically reserved for the highest animals; principally in humans, this includes consciousness. System 2 emerges from various structures of the forebrain with the greatest complexity originating in the cerebral cortex, and perhaps maximally from the frontal cortex. Connectivity between the frontal cortices, particularly regions within the prefrontal cortex, is very slow to develop. (Dosenbach et al. 2010; Fair et al. 2007, 2009). This is the seat of executive function and likely a major, but not the only area for agency and consciousness. Recall that Klemm suggests that agency and consciousness are not necessarily centralized to one structure or network. There are likely multiple areas involved in agency and consciousness, particularly including the orbitofrontal cortex and amygdala, centers central for decision making, emotional regulation, motivation, and personality (Depue and Collins 1999; Kalivas and Volkow 2005; Sanfey et al. 2003).

Toates' theoretical construct can be combined with Klemm's propositions to provide directions for future exploration into the neural correlates of agency. More specifically, differential examination of System 1 and 2 behaviors, particularly those related to executive functions, provides a possible starting point to continue the search for the agency conatogram. Starting in the dorsolateral prefrontal, parietal, and cingulate cortices, Klemm's ideas suggest that CIPs should be the target of interest and that the pattern of activity over time, and not the behavioral outcomes specifically, should be the focus of the search by neuroscientists. Neural activity over time should be the focus because decisions are not point events and we should not expect that the neural correlates will be point events either. As highlighted by Klemm, decisions are guided by one of two properties of the nervous system: (1) winner-take-all circuitry or (2) guide-gating mechanisms. Both of these involve a mix of interactions and competitions (e.g. excitation and inhibition) among relevant neural populations. Both of these possibilities are mediated by action and inhibition between forebrain regions (particularly within association cortices) and moderated by the neural activity of networks for memory, motivation, drives, and sensory stimulation.

This all suggests that each of the moderating networks can also influence the higher level networks that generate agency. Klemm suggests that intentions and decisions require a specific mechanism that can orchestrate the final behavior among these disparate neural subnetworks. He proposes that local oscillating activity by populations of neurons provides the main mechanism for the agency network. In effect, he is stating that there is not just one CIP or conatogram, there may be many, and the nature of the final behavioral outcome in these depends on the "neighborhood." Synchronized oscillations, neighbors working together so to speak, are what produce the intent and decision making that comprise agency in humans. If memory

networks also oscillate in a supporting fashion, then intent and a final decision to act are fully supported. The result and behavioral outcome is ultimately produced. This means that the firing of an individual "deterministic" neuron is largely inconsequential, and that agency and consciousness only occur through the action potentials of large populations of neurons working or not working (i.e. through collective inhibition) in synchrony. These population networks all get a "vote" to varying degrees. What makes a given vote stronger or weaker at producing neural oscillations exactly is unclear. However, it is clear that these oscillations can be generated either externally or self-initiated by existing self-referential networks. This last point presents a target for further exploration for identifying the possible networks involved in agency, particularly for consciousness.

9.4 Conscious and Unconscious Elements of Agency

The deterministic nature of neuronal physiology tends to trap neuroscientists and other psychologists. Each neuron discharges in an all-or-nothing fashion leading some to believe that the deterministic processes of neurons mean that agency is also deterministic. As is evidenced in the historical review of agency research, such nonlinear deterministic processes are generally held in contempt potentially because they suggest that there is very little ability to modify neural patterns within these processes. Clearly, this is a possible point of contention within and across various disciplines. Within the philosophical tradition, it is generally believed that "agency" cannot be unconscious. In this volume, Schlosser states that agency is not simply action or basic motor behavior. He argues that agency does not include "automatic" physiological activities of the body. From his perspective, the key aspect of agency is intent, which is a conscious process. However, it is not clear from any of the presented perspectives or recent empirical evidence whether a conscious process can be the initiator of the unconscious elements of agency, some of which are deterministic. This too provides a viable target for further examination in both philosophy and psychology.

While contentious, this pattern of events suggests that agency is still an outcome of a deterministic process. Klemm argues that while the final behavioral outcome may not be determined, ultimately "agency" is an unconscious process for all animals. As presented, patterns of neural activity among different networks lead to relatively predictable behaviors, behaviors we generally see as habits or fixed-action patterns. For most behavior, this collection of competitive network activity and synchronous oscillation generates the basic behaviors that ultimately produce complex behavior—in effect, this may be the ghost in the machine. That is to say that agency is largely unconscious and diffusely located across the central nervous system.

Across commentators, however, there is disagreement regarding the "agency" of automatic behaviors, behavior produced by System 1, or habits. The point of disagreement relates to how the behaviors are produced and whether they can even represent agency. Regardless of whether agency includes both conscious and un-

conscious behaviors, all of the contributors are clear that consciousness is a necessary condition where agency can be observed. This highlights a potential inconsistency regarding the definition of agency across science, or it may suggest that agency processes change over time. It may be possible that when a new thought or behavior occurs the first time, it is processed in System 2, but through continued practice it can transition to System 1. Put another way for humans, it is possible that new behaviors and thoughts require conscious agency, but eventually the activities that once required agency may become unconscious. This accounts for some of the differences in the use of the terminology, but it cannot be evaluated sufficiently until we identify the CIPs or conatograms that comprise agency in the first place.

Alternatively, Zhu suggests that regardless of the behavior, consciousness is involved to some level. He contends that all behaviors must be consciously derived because they require attention and purposeful action. However, Zhu points out the inherent confound of all neuroscience and psychology research on the topic of agency is that it is always conducted in conscious and awake individuals, and not in those in an altered state of consciousness (e.g. sleep, coma, anesthesia, hypnosis, etc.). Thus, he suggests that until we conduct agency research in unconscious behaving humans, consciousness and agency cannot be separated. Taking these points into consideration, psychologists should be careful to delineate when they are speaking about unconscious and conscious elements of agency.

According to Klemm, consciousness is the brain's observer of unconscious activity, but it is not a passive observer. This is where his ideas differ greatly with the broader field of neuroscience, but they are not entirely foreign as it is broadly accepted that changes in thoughts can produce changes in various biological mechanisms. Schlosser notes that this is a compatibilistic approach that may converge with philosophical perspectives. This idea is important because Klemm sees consciousness as a discrete set of CIPs—"an avatar to act on its behalf." Most neuroscientists would likely state that if there is an avatar or discrete set of CIPs, they are unable to take action, they are unable to be an agent of the body, and that consciousness is only an observer. In other words, the entire system is deterministic in nature and humans really have no agency or freedom to act as they wish. However, Klemm argues that we have two systems: An efficient system for unconscious activity that occurs through deterministic processes and another that is resource intensive, slow, and focused on neural processing of conscious, deliberate, analytical, and novel experiences that is not deterministic.

While behavior is the result of these processes of agency and consciousness, they also provide stimuli that can alter the system that generates agency. An interesting aspect of this proposal is that Klemm suggests that higher animals can modify CIPs and also modify ensembles of neurons that generate the CIPs. This idea was somewhat foreign a few decades ago, but we know that structural and functional elements of even the adult mammalian brain readily change to changes in behavior, training, the environment, or experience in a myriad of ways (see Draganski and May 2008; Holtmaat and Svoboda 2009; Jenkins et al. 1990; May 2011 for review). This evidence serves to support the idea that consciousness is not deterministic and potentially that it can modify agency as defined by Klemm.

This generally supports Toates' perspectives on "goal-directed behavior" within Systems 1 and 2. Within his perspective, conscious agency is defined by goals, intention, and flexibility of behavior in achieving those goals. Toates could be seen as hedging this perspective a bit when he focuses on incentives or reinforcers and on the neurotransmitters likely involved in modifying agency. For strict behaviorists, reinforcers, stimulus–stimulus relationships, and stimulus–response relationships all suggest a deterministic process at the simplest levels of behavior.

Likewise, for many neuroscientists, neurotransmitters are deterministic in that many open ion channels in individual neurons in a determined manner. However, Toates makes an important point that all of these systems can be altered, modified, and even hijacked through modification of neurochemicals in the body. Dopamine is a popular target as a mediator of consciousness, or at least of the major elements of consciousness (Palmiter 2011). Acetylcholine is likewise another target for a correlate of consciousness and agency due to its pervasive presence in the brain (Perry et al. 1999; Woolf and Butcher 2011). These are worthy targets for seeking the neural correlates of agency because their presence may provide the breadcrumbs for neuroscientists to follow within systems. However, it seems more likely that multiple neurotransmitters and neuromodulators will be colocated across numerous circuits and systems making them relatively poor targets for the final pathway(s) of agency and consciousness. Klemm's CIPs and interactions among systems of systems within the brain are likely better targets for locating the centers of agency in animals.

In the end, Toates considers Klemm's proposals through identity theory and he ultimately appears to emerge on the same side as Klemm in that humans can consciously modify their intent. That is, a conscious intention and goal-directed behavior can alter conscious and unconscious processes, including many that are deterministic once initiated. Klemm then states that the CIPs involved do not cause consciousness, they are consciousness, and therefore are the best target for further research.

In this context, philosophers like Schlosser and Zhu would likely contend that only Toates's System 2 truly can involve agency due to the relative role of consciousness and intent. More specifically, Schlosser and Zhu suggest that neuroscience and psychology should focus on "reasons," the process of reasoning, "intention," and executive functions, which leads to a focus on conscious decision-making behavior. It is likely in the neural correlates supporting these behaviors where we will find the agency conatogram.

The question of consciousness, therefore, becomes a central focus of agency across disciplines. Klemm contends that elements of agency are evident from basic organisms at one end of the spectrum to humans at the other. This evolutionary perspective is logical especially considering that the neural structures continue to build from the most basic neural network to exceedingly complex networks on top of networks. Klemm contends that in these structures of networks there is a point, albeit as yet undefined, where self-referential networks produce self-awareness. The commentators seem to agree that consciousness only emerges through the complexity of the neural substrates that comprise organisms—most of which focus on humans as higher animals.

This brings us to the final topic of free will, a topic that is addressed with some reluctance by all commentators. Among the commentators to this volume, and

seemingly science more broadly, there is concern about addressing the topic of "Free Will." Where once it seemed that consciousness was untouchable academically, now it seems that free will has become a bit of a third rail for many academics. Concerns over metaphysical, supernatural, and paranormal views were all presented as reasons to tread lightly on this topic. As noted by Schlosser, it would seem that a large part of this concern comes directly from developing an operational definition of free will. As repeatedly highlighted by nearly all contributors, we cannot seem to agree on the operational definition of agency, a much less contentious topic. Therefore, the starting point for free will is perhaps too uncertain at this time for meaningful scientific discourse. Likewise, perhaps the response to Libet's research and his reaction also reveals why this is such a contentious issue. From this, Schlosser makes a suggestion worth consideration that as scientists we should be very clear to operationally define and differentiate agency and free will. A clearer conception of the two, especially in the context of conscious versus the unconscious, will help interested scientists from diverse disciplines evaluate the nature of the research and empirical evidence that is produced. While this may seem elementary, it is interesting how many contributors to this volume raised the issue. This alone is evidence that the topic requires further collective consideration.

More broadly, setting aside the clear concern over free will as a focused topic, it is instructive to consider the implications of these perspectives on other topics included by commentators to this section and larger questions of agency beyond the individual. As is always the case, individuals do not exist alone within a vacuum and their neurobiology will have an impact on the individual, their behavior, their environment, and even broader sociocultural issues. Klemm addresses the point that agency is not simply a matter for academic consideration; it is instead "a very practical matter in the real world of human behavior." This is where the topic of free will might be academically addressed, albeit indirectly, in a functional and safe manner mostly around an operational definition employing exemplars of "conscious decision making."

As beings that act in the world, we often look around for explanations and predictions of how or why humans behave in a particular way. We are quick to state that once we understand how behavior works, we will use it to accept, ignore, or excuse various behaviors. In other words, the understanding can remove an individual's sense of agency and related responsibility, at least in humans. Toates does well to address this topic by addressing the question of violence and responsibility. He importantly notes that the legal profession is quick to adopt developments in neuroscience and psychology as reasons for humans to act without agency. This matter is more fully considered in the latter part of this volume.

9.5 Conclusion

The purpose of this section was to address the neurobiology of key elements of agency in a comprehensive and axiomatic way. A search of phrase utilization (called "ngrams") in the Google Ngram Viewer from the roughly 5.2 million books in-

cluded in the Google Books archive shows that from roughly 1980 to 2000, use and growth of the terms "human agency" and "neuroscience" track tightly suggesting an association between the two ("Google Ngram Viewer," n.d.). Accordingly, this section of the volume was intended to integrate research on these areas and presents their relative value in the context of other topics related to agency, including the history and broader philosophical considerations.

The empirical and theoretical considerations addressed suggest that "agency" is not a singular concept. The difficulty in describing, defining, and researching this topic and its neural correlates clearly suggests that agency must be subdivided into at least two areas for further exploration: conscious agency and unconscious agency. For those that contend that this is a uniquely human capacity, then we can use Toates' construct to examine agency within Systems 1 and 2, respectively, to organize the relative inputs and outputs.

On the whole, the question remains regarding the neurobiology of agency. Is there a "conatogram" or "conatograms" that account for agency and consciousness? Alternatively, is there a ghost in the machine? Decidedly, research is required to further answer these questions. The ideas presented by the four commentators provide a clear point of departure for further consideration and research into the neural correlates of agency. At least among this collection of scientists, there is a sense that agency and consciousness are very likely centered and diffusely distributed within and across the central nervous system, mostly in the brain. However, the exact location of the CIPs or conatograms that comprise all related elements of agency will require novel research approaches to parse out the relevant and specific components involved.

In conclusion, at least among psychologists who are interested in neurobiology, neurophysiology, and the link to behavior, it is clear that agency provides a very useful concept for consideration of behavior produced by animals. This section also reveals that at least at this time, the concept of free will is not useful in most academic pursuits at this time. I would like to echo the thoughts of Klemm from his response to the commentators in that hopefully we can use his axioms and propositions for further critical analysis and research into the topic of human agency.

References

Baillieux, H., De Smet, H. J., Dobbeleir, A., Paquier, P. F., De Deyn, P. P., & Mariën, P. (2010). Cognitive and affective disturbances following focal cerebellar damage in adults: A neuropsychological and SPECT study. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 46*(7), 869–879. doi:10.1016/j.cortex.2009.09.002.

Bidney, D. (1962). *The psychology and ethics of Spinoza: A study in history and logic of ideas*. New York: Russell & Russell.

D'Angelo, E., & Casali, S. (2012). Seeking a unified framework for cerebellar function and dysfunction: From circuit operations to cognition. *Frontiers in Neural Circuits, 6*(January), 116. doi:10.3389/fncir.2012.00116.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *The Behavioral and Brain Sciences, 22*(3), 491–517; discussion 518–569. http://www.ncbi.nlm.nih.gov/pubmed/11301519.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. doi:10.1146/annurev-psych-113011-143750.

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., & Schlaggar, B. L. (2010). Prediction of individual brain maturity using fMRI. *Science, 329*(5997), 1358–1361. doi:10.1126/science.1194144.

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioural Brain Research, 192*(1), 137–142. doi:10.1016/j.bbr.2008.02.015.

Fair, D. A., Dosenbach, N. U. F., Church, J. A., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., & Schlaggar, B. L. (2007). Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences of the United States of America, 104*(33), 13507–13512. doi:10.1073/pnas.0705843104.

Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., & Petersen, S. E. (2009). Functional brain networks develop from a "local to distributed" organization. *PLoS Computational Biology, 5*(5), e1000381. doi:10.1371/journal.pcbi.1000381.

Google Ngram Viewer. (n.d.). https://books.google.com/ngrams/graph?content=human+agency%2C+neuroscience&year_start=1800&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Chumanagency%3B%2Cc0%3B.t1%3B%2Cneuroscience%3B%2Cc0. Accessed June 30, 2014

Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience, 10*(9), 647–658. doi:10.1038/nrn2699.

Jenkins, W. M., Merzenich, M. M., & Recanzone, G. (1990). Neocortical representational dynamics in adult primates: Implications for neuropsychology. *Neuropsychologia, 28*(6), 573–584. doi:10.1016/0028–3932(90)90035-M.

Josselyn, S. A. (2010). Continuing the search for the engram: Examining the mechanism of fear memories. *Journal of Psychiatry and Neuroscience*: *JPN, 35*(4), 221–228. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2895151&tool=pmcentrez&rendertype=abstract.

Kalivas, P. W., & Volkow, N. D. (2005). The neural basis of addiction: A pathology of motivation and choice. *American Journal of Psychiatry, 162,* 1403–1413.

Koch, C., & Crick, F. (2001). The zombie within. *Nature, 411*, 893. doi:10.1038/35082161.

Lashley, K. (1950). In search of the engram. In experimental biology symposium No. 4: Physiological mechanisms in animal behaviour (pp. 454–482). Cambridge: Cambridge University Press. doi:10.1097/00008877–199204001-00015.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain, 106*(3), 623–642. doi:10.1093/brain/106.3.623.

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences, 15*(10), 475–482. doi:10.1016/j.tics.2011.08.002.

Palmiter, R. D. (2011). Dopamine signaling as a neural correlate of consciousness. *Neuroscience, 198,* 213–220. doi:10.1016/j.neuroscience.2011.06.089.

Perry, E., Walker, M., Grace, J., & Perry, R. (1999). Acetylcholine in mind: A neurotransmitter correlate of consciousness? *Trends in Neurosciences, 22*(6), 273–280. http://www.ncbi.nlm.nih.gov/pubmed/10354606.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*(5626), 1755–1758. doi:10.1126/science.1082976.

Schmahmann, J. D. (2010). The role of the cerebellum in cognition and emotion: Personal reflections since 1982 on the dysmetria of thought hypothesis, and its historical evolution from theory to therapy. *Neuropsychology Review, 20*(3), 236–260. doi:10.1007/s11065–010-9142-x.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*(5), 543–545. doi:10.1038/nn.2112.

Thompson, R. F. (2005). In search of memory traces. *Annual Review of Psychology, 56,* 1–23. doi:10.1146/annurev.psych.56.091103.070239.

Woolf, N. J., & Butcher, L. L. (2011). Cholinergic systems mediate action from movement to higher consciousness. *Behavioural Brain Research, 221*(2), 488–498. doi:10.1016/j.bbr.2009.12.046.

Part III
Looking at agency from the top

Chapter 10
The Relational Basis of Agency: An Integrated Psychological/Theological Approach

Philip Browning Helsel

Agency has become an important theme in recent psychological literature, and debates have sparked over the source of agency and the capacity for agency in the face of seemingly deterministic social limits. In this chapter, I examine how agency fostered across the life span should be a *shared* agency by exploring a theology of agency.

Agency—the feeling that one can do something to influence the world[1]—arises from experiences of being seen and of witnessing one's own actions mirrored in another's response (Ellison 2013). First, God is known most fully in relationships in which people exert their agency on behalf of each other's flourishing: indeed, God *is* the care exhibited in such relationships. Second, consistent empathy, in which God is revealed, becomes the ground out of which agency is formed. Using recent developments in attachment theory and neurobiology, I will argue that agency is indirect since it comes through the experience of consistent mirroring. Agentic relationships both reflect God's presence and impact God.

To claim that agency is important is not to claim that it is of central importance. Indeed, agency seems important to those who are not able to express agency directly or who have their agency circumvented. Some of the most significant contributions to the theological discussion of agency come from feminist theologians, frequently responding to issues of trauma and violence against women, who are attentive to the diminished agency of women in many cultures (Beste 2007; Hoeft 2009; Suchocki 1994).

The capacity to witness oneself acting with agency is not equally available across the life span: Young children and older adults have less agency than middle-aged adults in many societies. In this sense, young children and older adults often need

[1] The primary difference between *agency* and free will is that in the psychological literature agency is often linked with the capacity to develop and attend to goals, whereas the concept of free will is often posited as a philosophical or even ontological issue, having to do with the capacities or the traits of the person with a spirit or soul.

P. B. Helsel (✉)
Boston College, Chestnut Hill, MA, USA
e-mail: philip.browninghelsel@bc.edu

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_10

advocates with more agency. Persons with less agency, at any given time, depend upon others, leading to the conclusion that agency must be shared with those who have the least agency at any given point (see Fig. 10.1).

Fundamentally, agency is not the only "good" and not the sole "virtue" that is significant to human life. Other important goods include community and connection, acting in a trustworthy manner, and the capacity for reasonable self-sacrifice.

Nevertheless, shared agency should be included among the "goods" that are available to persons since the capacity to act with agency is, at least for Christians, an important part of their capacity to participate in God's purposes. From a Christian theological standpoint, shared agency is rooted in the image of God although separable from it. Persons participate with God when they use their agency for the well-being of others, yet those who are no longer able to participate explicitly in God's purposes are still bearers of the image of God (Swinton 2012). In these cases, shared agency must be used by others to help those whose agency is diminished to fully participate in lives of service, worship, and vocation.

In psychology, there is a current controversy between what has been termed top-down and bottom-up agency (Prinz 2012). A top-down notion of agency assumes that people are able to develop goals and work towards them; bottom-up agency assumes that people are determined by factors beyond their control, operating within systems that constrict agency so severely that it renders agency nearly unintelligible.

In this chapter, I maintain that top-down agency is a possibility for some, but, paradoxically, it arises from a complex set of bottom-up interactions and is thereby fragile, contested, and malleable. Additionally, I argue that God is intimately connected to the human struggle for agency so that this discussion is not strictly a psychological discussion, but also has theological ramifications with God working alongside persons for the sake of their agency as they use their agency for the well-being of communities.

My own approach to agency appears to resemble bottom-up agency but is actually more complex than this: Persons act with agency based on early and ongoing experiences of empathy, but this capacity for agency can also surprise and overturn systems where empathy has not been accorded. This capacity for agency—when one would have expected none—arises itself from relationships where there have been signs of care and mirroring. Understanding the interpersonal formation and maintenance of agency through activities of empathy is thus crucial to fostering agency.

10.1 Shared Agency in Hebrew Scriptures and Christian Theology

Christian theology offers a distinctive vision of the human person known as theological anthropology. In theological anthropology, *agency* is frequently described with the term *freedom* or *free will* and is thereby at the center of a range of complex debates; nevertheless, there is a paradox in this accent on freedom. For many

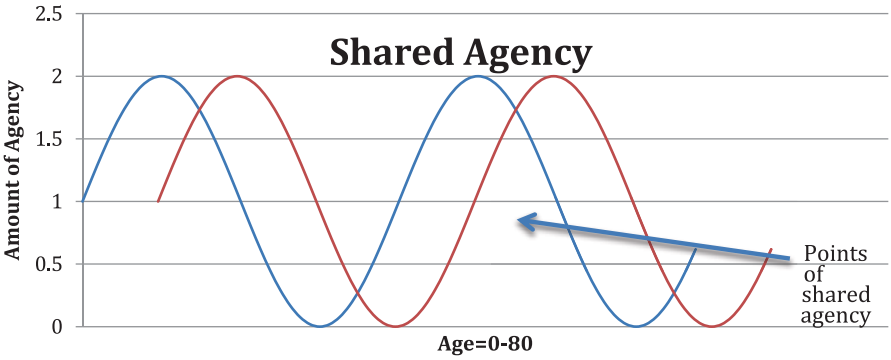

Fig. 10.1 Sharing of agency between person over life course

Christian theologians, freedom, or personal agency, is directed in service to others. In this succinct chapter, I will explore only Christian theology and leave other religious perspectives for others' analysis.

Drawing on the traditions of the Hebrew Scriptures, in which Israel was depicted as having agency apart from God, the agency of persons as a capacity has been seen to be a key constituent of the human person, but this was not agency for its own sake. Rigby (2001) indicates that in the Hebrew Bible text of the Psalms, brief poems or hymns sung in worship, "double agency" is at work in which both God and persons contribute to their deliverance from trouble. For the ancient Israelites, agency was fulfilled in worship to God and service to their neighbors, especially the ones who were marginalized.

Christian theologians have followed these Hebrew traditions of seeing human agency as most completely fulfilled in the service to God and to others, thereby accenting shared rather than personal agency. Catholic theologians, in the virtue tradition, have linked the concept of agency to the capacity to live out the virtues of faith, hope, and charity (Hollenbach 2002). Pope John Paul II, (1981) in his encyclical on meaningful work, suggested that personal agency was part of God's image in humanity:

> as the "image of God" he/she [Man] is a person, that is to say, a subjective being capable of acting in a planned and rational way, capable of deciding about him/herself, and with a tendency to self-realization. (*Laborem Exercens*, 6.2)

In this quote, the image of God in humankind is closely linked with agency. Nevertheless, in a Catholic framework, there quickly needs to be a statement about the direction towards which one's agency is aimed. Indeed, in the Catholic social justice teaching of *subsidiarity*, social relationships exist for the purpose of helping persons "in their free but obligatory" exercise of self-realization (Iber 2010). Catholic theology has traditionally claimed that God enables people to participate with God through *operative grace* (Stump 2010). Since grace adds to nature, in the thought of the influential Catholic theologian Thomas Aquinas, Catholic theologies

are more likely to emphasize a person's natural capacity for agency as a reflection of God's goodness, pointing to a God who wants a person to be a steward of God's good creation. From this standpoint, and in consonance with the Catholic social teaching, we can see that God wants agency to be used on behalf of those lacking it at any given point.

On the other hand, Reformation theologians highlighted the limitations of personal freedom and emphasized God's grace at each juncture. In the starkest Reformation polemics, agency nearly disappears. In this view, agency is preceded by God's initiative. Despite this emphasis, the Reformers were also concerned with the individual's volition. God first gave persons the free gift of faith through grace; and actions undertaken with personal agency, what these theologians called "works," are considered an outpouring of the life of service in gratitude for the gift of faith. Paradoxically, freedom is achieved as the result of willing service to God. As Reformer Martin Luther put it, "Insofar as [a person] is free [they] do no works, but insofar as [a person] is a servant, [they] do all kinds of works" (Luther 1970). Luther described the transformation of will into servitude before God, with the stress throughout on the unmerited gift of faith. My view of this leads to a certain paradox: Although the agent was free, the person was also deemed incapable of acting with freedom, bound as they were by sin.

Continuing the Reformation era emphasis on individual authority in matters of conscience, in modern times a distinctly personal framework began to dominate theological reflections on freedom. Neoorthodox theologians of the twentieth century, such as Emil Brunner, Karl Barth, Paul Tillich, Reinhold Niebuhr, drew from existentialist thinkers to describe a situation where personal agency was in the forefront. For these theologians, persons confronted a radical choice to make meaning of their lives before a God who desired to be in relationship with them (Chopp 2007). Each person was confronted with a unique choice—a choice that was not transferable to any other person or entity—of either using his or her freedom in relationship to God or rejecting freedom for the meaninglessness of an existence apart from God. Real choice was possible, and this condition could provoke anxiety. The emphasis among these thinkers was different, but in the foreground was an individual "subject" in a personal relationship to God. This distinctive angle left aside important social dimensions of agency, leaving the individual largely alone with God. This personalist framework still influences theological thinking.

A personalist framework for human freedom would not suffice in situations of extreme poverty and marginalization since it seems too individualistic. Liberation theologians from Latin America challenged the privatized framework of twentieth-century existentialist theology since they maintained that true agency proceeded from proper positioning in society: Persons become agents when they participate with God's "preferential option for the poor" (Gutiérrez 1973). A person was truly free only in shared agency when that person used their freedom in "solidarity" with the marginalized in the two-thirds world (Ashley and Metz 2007, p. 212). For too long, theology had been done from wealthy Western contexts that maintained social power and privilege (Johnson 1992; Ashley and Metz 2007; Gutierrez 1973). In my estimation, liberation theology succeeded, through a Marxist critique, in locating the

material conditions of the living world as the foreground for God's action, and this constitutes an important revision to a realm of inquiry that has been too otherworldly.

Feminist theologians continued the critique of neoorthodox theology by reclaiming agency within a particular framework: the distinctive and contested experience of fostering women's well-being and justice. In recent decades, feminist theologians have challenged the concept of free will as being too individualistic and only referring to men. Drawing from women's experience of embodiment, suffering through domestic violence or sexual assault, and social marginalization, these authors foregrounded the concept of personal agency as a vexed issue in situations of gender oppression (Hoeft 2009). Women had been denied agency, and this was an injustice, not an abstraction. Here, agency was as important as empowerment. Rather than abandoning one's will in order to regain it in Christ—as Luther put it—feminist theologians argued that women first needed an opportunity to *have* a self (Jones 2000). A crucial question among feminist theologians was "Whose agency are we fostering?" with the implication that we should be fostering the agency of those pushed to the margins. Feminist theologians highly prized agency but discovered it in a community inspired by feminist concern for the full equality of women, protection of their bodily integrity, and the wholeness of creation, all of which were seen as flowing from the Divine Spirit that inspired creative freedom. In my work on gender roles, I have argued that gender roles operate as "injunctive norms" that rely on the belief that the biological factors predispose a person to a certain experience of the world and that persons are sanctioned when they violate these expectations based on biology (Browning Helsel 2009). Feminist theology thus helpfully indicates how women have been denied agency based on arbitrary and socially constructed categories, such as gender and sexuality.

One of the most important recent developments in theology has been process thought, which, although densely philosophical at times, actually has quite promising practical contributions to pose to the question of agency. Process theology attempted to address the theological difficulties posed by an image of an all-powerful God—often described with monarchical metaphors—and proposed that God is immanent to creation and working within it. In the mid-twentieth century, process theologians, influenced by the work of mathematician Alfred North Whitehead, argued that God should be understood as that force *within* creation that influences all creation towards becoming (Hartshorne 1948). Within a process framework, agency is important but from the standpoint of relational experience. Rather than individuals acting with personal freedom (yet in a process framework such freedom inevitably existed, even in the most extreme situations), there were *webs of being* through which people reached freedom through *connected acts of becoming*. Since God was not seen as omnipotent but described more as operating within creation as a force for good, the traditional problem of suffering—*How could a good God and all-powerful God allow bad things to happen?*—receded, and in its place there was an emphasis on cooperation with the *processes towards goodness* that God influenced into being (Hartshorne 1984). Process thought showed how natural ecosystems were linked to human systems and how persons existed in webs of experience even before their birth. A person's agency came through analyzing and changing

these webs of relationships. Rejecting traditional metaphysics that posited isolated individuals and an all-powerful God, process theology depicted an interconnected and interpersonal reality through which God influenced persons.

A process theological framework is among the most promising theological perspectives through which to view agency because it shows how God works fruitfully alongside creation, assisting a person's own freedom to lead to well-being. Within a process framework, the universe is truly free to respond to God's initiative. Yet, this agency is appropriately complex because it is situated in webs of social and natural relationships. A critique of process theology that suggests that it does not accord God enough agency has a place. Even if a person lacks relational mirroring in life, it is possible that God may act with agency in relationship to that person in order to foster their sense of personal agency. Nevertheless, process theology helps us understand the theological and social dimensions of the relational development of agency.

Several key insights emerge from this exploration of agency in Christian theology and in Hebrew Scriptures. First, freedom and agency have consistently been affirmed by Hebrew Scriptures and Christian theologians but with a *telos*—or goal—in mind. The capacity for agency reflects God's image, but persons do not always have equal agency. If personal agency is diminished in one's life span—as a result of illness, debility, violence, or misfortune—this does not mean that a person is *less human* or less of a person, or reflects less of the image of God, as a result. Rather, persons require shared agency with others who will empower their own expressions of agency.

In theological terms, it matters a great deal what use agency is put towards. In contrast to Enlightenment thinkers who praised autonomy for its own sake, Christian theologians have consistently valued freedom *in relationship* to others, namely, in relationship to God and to other human beings. Catholic theologians maintained that people were "naturally" able to respond to God when empowered by God's operative grace, and Protestant theologians implied agency as an aspect of the turning of the will over to God in service based on God's initiative of relationship. While neoorthodox theologians assumed a sense of agency in relationship to a personal God, liberation, feminist, and process theologians have affirmed the inherently social nature of agency. Freedom is meant to link a person to others in relationships of justice, challenging oppressive systems. In each case, agency has been an important issue. Recent developments in theology are especially promising because they suggest that agency is not only effective in the interpersonal sphere but also has an impact on God.

10.2 A Constructive Contribution to a Pastoral Theology of Agency

God operates to engage with and transform a person's life by using human agency. At the end of a developmental process through which agency is fostered, God can be seen as the force that makes agency possible. Given our exploration of agency

within the Hebrew Scriptures and Christian theology, agency emerges as a real but limited ethical good; it is most salient when it is expressed as shared agency devoted to both the good of oneself and the good of the community. It does not define one's worth or dignity: The amount of agency a person has may vary across one's life span (see Fig. 10.1). Nevertheless, agency is an important good since persons are meant to participate with God in the establishment of a just and caring society and in the stewardship of God's creation.

Agency is a shared interpersonal good rather than an individual achievement. Agency does not imply autonomy but is the effect of a network of relationships. A pastoral theological approach to agency addresses some of the conflicts in psychology about whether agency is the personal capacity, almost like a possession, of someone able to achieve goals (top-down agency) or is largely subject to deterministic social forces (bottom-up agency). As I suggest, top-down agency emerges through a series of interactions with bottom-up forces, confirmed and realized in relationships in which agency is affirmed or denied. Shared agency, the responsibility to use agency for the fostering of those who have been denied agency or who lack agency at any given point, is an important social responsibility (see Fig. 10.1).

This chapter also seeks to make theological claims: God wants persons to be able to balance agency with a variety of other goods and experience more agency at certain times in the life span; nevertheless, the systematic denial of agency for some is a denial of the goodness that God wishes for the human community. Using a process theological approach to agency will help explore how, if people are systematically denied agency, this violates God. When a person honors another's agency and fosters a society where agency is shared, this contributes to God's well-being. In other words, it is necessary to offer an ethical framework—here rooted in feminist process theology—in which the goals of agency are more clearly defined. Otherwise, agency risks becoming a somewhat empty conceptual category. This attention to agency should be fostered especially on behalf of those whose agency is not likely to be considered. The well-being of the world and God's well-being are intimately connected in a process framework.

10.3 Process Theology and Relational Agency: Why a Violation of Another's Agency Constitutes a Violation of God

In this section, I will explore one of the most promising recent arguments for how God is connected with human agency. Using this resource will help me explain how violations against the agency of persons constitute a violation of God. Given our review of Christian theology in the first section, this approach is situated within the feminist process theology. Feminist process theologians have challenged classical notions of sin, arguing that sin is interpersonal and it is environmental, considerably broadening the individualistic focus of theology. In what follows, I offer a detailed description of agency in the thought of a recent theologian in order to explain how agency impacts God.

Theologian Marjorie Suchocki (1994), in her book *Fall to Violence: Original Sin in Process Thought* revises the concept of original sin by showing how sin is rooted in interpersonal violence. Original sin is the classic notion that, through the Fall of Adam and Eve in the Garden of Eden, persons have been in a state of sin, alienated from God, and are in need of redemption.

She critiques and builds upon the example of Reinhold Niebuhr, a mid-twentieth-century neoorthodox theologian, who maintained that persons were capable of "self-transcendence." For Niebuhr, persons were both "Nature" and "Spirit," that is, they both belonged to the created order as creatures and were able to reflect on their existence, even the totality of their existence. "Spirit" was the capacity for reflection, and this ability to review the totality of one's life led to existential anxiety—questions about the meaning of life and fear of death—which led, in turn, to a desire to repress that anxiety. Niebuhr maintained that once persons felt the existential anxiety of being a limited person, they chose through a variety of means to secure themselves against this anxiety, and the chief of these means was pride. In Niebuhr's formulation, pride was "the desire to be like God," and Niebuhr, following an Augustinian line of thought, saw pride as a primary sin from which all other sins, such as violence, flowed. Securing oneself against anxiety by pride was the basic sin for Niebuhr (2004). As we saw from a feminist perspective, it was necessary to "have" a self before giving that self away; the sin of pride is but one form of sinfulness along a continuum of self-forgetfulness and self-sacrifice.

Suchocki objected to Niebuhr's individualistic rendering of sin as pride since it downplays the systemic and structural violence towards which women are routinely subjected. Rather than pride, she maintains that violence against the well-being of others is the primary sin. She maintains that a kind of interpersonal violence surrounds persons before their birth and thus could be considered a form of "original sin," in that it is a form of violation of others that precedes personal responsibility. Fundamentally, the conditions that foster violence against others are well established before persons enter into them socially as fully cognizant members of society. For example, a person is born into a particular social identity that confers privilege or marginalization—at times based on external appearances—without choice. In this context, agency becomes an important act of resistance against the pressure for social conformity or identity expectations, an opportunity to use one's social role more flexibly and challenge the conditions that lead to marginalization.

The systemic nature of violence makes it difficult to notice and resist, but it is nevertheless real in its effects. As "rebellion against the well-being of the world," stripping away the agency of others, violence renders some incapable of truly thriving. Suchocki argues that people are formed in relationships that foster interpersonal violence and violence against the Earth (Suchocki 1994, p. 60). Just as the notion of agency does not develop as the innate property of a particular individual but rather within a set of relationships, theological notions of sin and salvation that focus on an individual's relationship to God alone are inadequate because they do not grapple with the real harm that persons do to each other and to their environments. Suchocki invites us to broaden our notions of agency to include interactions with the natural

world, as well as social world, in order to understand how these relationships, and the patterns through which they are negotiated, can be severely harmful.

She argues that traditional doctrines of original sin are not relational enough: They tend to emphasize an isolated God and isolated individuals. By contrast, God exerts agency *in relationship* with creation, influencing creation towards "interrelated communities of well-being" (Suchocki 1994, p. 60). God is not separate from creation, influencing it from the outside, but continues to care for creation through God's influence. At the same time, God is dependent on the agency of creation to respond to the concerns of a suffering world. Profoundly empathic, God feels the experiences of the world along with the world so that agency is more squarely set within the framework of creation rather than on God's side, acting outside or beyond creation.[2] From this perspective, God is the one who feels, along with creation, all the suffering that creatures endure.

Original sin is that disruption of God's communion of care that involves "unnecessary violence." It precedes persons so that they are born into communities that foster such violence and are educated into it and yet they are also capable of resisting it through cooperation with God towards well-being (Suchocki 1994, p. 57–58). Since God is intimately connected to the world's well-being and the well-being of all those in the created order, God's well-being is tied to human agency. At the same time, God influences agency and cooperates with it without taking it over, which thereby gives agency to individuals. Given a process theological framework, God instantiates new possibilities for action in a person's life but does not control the outcome. "The integrity of the world's self-creativity in response to the creativity of God is the real freedom, to whatever degree, of how it responds to all the forces, including God, that impinge upon it" (Suchocki 1994, p. 58).

Suchocki argues that freedom is contested and contextual yet is real. Persons have a "response-ability" that "is at the core of every moment of [their] lives" (1994, p. 132). By this, Suchocki means that we are meant for relationships with God, humanity, and creation; my argument extends her insights further to suggest that relationality itself is a gift that comes, paradoxically, from first being in relationship with others. In process theology, freedom is real and constrained so that the exercise of one's freedom is a central task that has an impact on others and on God.

Nevertheless, this response-ability is not merely given; it is structured in networks of power that significantly shape the range and scope of this response. "Freedom is indeed conditioned … If [freedom were not limited] relationality would be

[2] Suchocki (1994) does believe that redemption of structural sin happens through Christian faith and argues that Christianity provides a specific saving knowledge of God. She maintains that persons can know God through Jesus Christ, but this is particular knowledge rather than universal knowledge: It is known through a certain kind of "perspective" (p. 53). Since it argues that God became a person in a particular time and place, Christianity affirms the creation as the space in which God exists. On the other hand, God's perspective is broader than God's incarnation in Jesus Christ, so that God's knowledge encompasses all cultures and religions rather than being limited to one, allowing the provisional revelation of the Christian faith to be experienced in a variety of contexts without eliding the distinctiveness of other forms of knowledge. Suchocki maintains that it is through incarnation that God is linked to creation, moving with agency towards the well-being of all.

meaningless. One does not have a set amount of pure freedom … rather, the influx of relation that goes into our moment-by-moment creation forces a response that is yet paradoxically free" (Suchocki 1994, p. 132). Suchocki shows how freedom is real even while it is limited. Freedom is fragile, tenuous, and relational, depending on both human action and divine intervention. The agency that a person exerts does not occur outside of social constraints or embodiment, but occurs within the web of relationships in which a person finds themselves.

In Suchocki's feminist process approach, God is involved in the ongoing progression of relationships between human beings, creation, and Godself. Ethically, this means that people are meant to foster the development of agency between each other and to violate this agency amounts to a violation of God. God's possibilities are tied up in the relational possibilities of creation, even while God continues to influence creation towards greater well-being. For good or ill, a person's actions make an impact: "Apart from a response to relation, there is no coordination of relation, and hence no coming to be of the relational reality" (Suchocki 1994, p. 132). Agency is found in and through relationship—it is difficult to separate out the agent as a "subject" that acts on "objects" outside the self.

In process theology, God is not over/against creation but rather within/alongside creation, so that God is that force that allows creation to reach its full potential. Indeed, in process theology, God can be identified with relationality—that energy that exists when communities are working together for the well-being of the earth and for the flourishing of themselves and other communities. While process theology gives you a sense of God's intimate closeness with human agency, it lacks the depiction of God as a separate being towards which one could direct one's prayers (Phillips 1965–2014).

How does this look on a practical level? Putting process theology to work in the context of this chapter, I maintain that we should explain God as existing in the nexus between creatures rather than outside of them as an overpowering alien force. God assists persons to use their agency to cooperate with others for the good of creation and indeed, in a process sense, God *is* this spirit of cooperation existing in the connections between persons. In other frameworks, this could be described through *pneumatology*, or the language of the "Holy Spirit," the third person of the Trinity. The consequence of this argument theologically is that when people violate each other's agency, they violate God, who exists at the nexus between creatures, in the webs of relationality in which they operate.

How does this change our view of agency? Suchocki argues that transcendence needs to be revised in a horizontal direction. If a person's sense of self, including the development of agency, is linked with all other creatures in a process of becoming, this relational emphasis seems to be justified. We will explore this point further in the next section. She argues that the infinite is not found beyond time, but is experienced in the progression of time and in the context of the natural world. Therefore, "If infinity is no stranger to nature, then one can develop a 'horizontal' notion of self-transcendence, such that it is gained through a certain 'with-ness in' the world, not an 'over and above' the world" (Suchocki 1994, p. 35–36). She describes these

capacities for horizontal self-transcendence as pertaining directly to the experience of time, being manifested in "memory, empathy, and imagination" (1994, p. 95).

Although she does not make this claim, in my own view as a pastoral theologian, I see these three aspects—memory, empathy, and imagination—as logically ordered. If the past is dominated by violence, then empathy will become difficult and imagination dangerous, if not impossible (Ellison 2013). While empathy provides a powerful capacity for connection, people cannot always realize this connection because of violent self- or other-experiences that shape them. If transcendence can be experienced in the relationality of human persons in these three areas, then it can also be lost or violated in each of these areas.

While persons seem capable of self-transcendence through memory, empathy, and imagination, they are also born into webs of relationships that allow and condone violence against others and against the earth. This has a central impact on one's experience of time of oneself as a person with a past, present, and future. "One can roughly correlate a failure of transcendence through memory with the perpetuation of the past as violence; the failure of empathy has a correlation with violation mediated through the solidarity of the human race; and the failure of imagination relates to violence perpetrated through social structures" (Suchocki 1994, p. 36). Not naïve optimism, a process viewpoint shows the enduring and embedded nature of structural violence because it highlights how actions undertaken in a web of being spread out to influence others.

Shifting the focus from the vertical-God relationships to the horizontal-human relationships allows agency to emerge as a theological concern in horizontal self-transcendence. Suchocki argues that the relationality through which response-ability develops is by no means a given. Through relationships of empathy early in life, and through the fostering of such relationships across the life cycle, people can experience freedom and imagination; they sense the promise of the future. Each relationship contributes to a sense of agency, those with early caregivers, the created world, and God. Since God exists in the nexus between all these relationships, each person's sense of agency impacts God.

For Suchocki, agency is a central theological category that not only makes guilt and sin possible but also makes freedom and liberation possible; nevertheless, it is essential that this capacity or freedom not be understood in autonomous terms, but be described as inherently relational, since human persons are necessarily connected in webs of relationship to others, to the natural environment, and to God. While each of the modes of self-transcendence through relationship can be violated, it is essentially some form of agency that allows a person to challenge this violence and redeem the situation. "The ontology of relationships maintains that the very possibility of relationships depends upon the ability to respond to relationships, and that this 'response-ability' is at the core of every moment of our lives" (Suchocki 1994, p. 132). Freedom is real.

Although she critiques Niebuhr's notion of "nature" and "spirit," Suchocki allows for the capacity of self-reflection, a "spirit," though which one, is able to evaluate the consequences of one's actions. She considerably broadens the notion of sin beyond Niebuhr's narrow view of pride and she also envisions persons as

responsible for sin and capable of change, able to exercise memory, empathy, and imagination on behalf of others. Response-ability is no mere academic exercise for Suchocki; it is vital to the flourishing of oneself, others, and creation.

Since God is not an alien force or distant power, but is seen precisely where people use cooperation with each other to foster well-being, the violation of a person's agency constitutes a violation of God. As we have seen, agency is real in a process framework, and this means that God does not control the outcome of human relationships, but rather influences them and activates their possibilities. From whence does the capacity for self-transcendence come? According to Suchocki, it comes from one's experiences within the web of relationality—despite conditions of violence, there may be a sign that provides a basis for agentic action on behalf of memory, empathy, and imagination.

Next, I will explore the developmental progression that allows for personal agency and show how God exists in the connections between persons as agentic relationships are formed. This requires bringing to the foreground some of the promising research at the intersection between attachment theory and neurology. This research confirms and extends Suchocki's thesis. We must revise personal or self-contained notions of individuality and perceive the horizontal dimension of relationships: Agency is shaped in relational forms through practices of other witness, and for this reason, the development of agency is a theological and interpersonal concern. If, as process theology suggests, God exists at the linkage points between individuals, the interactions between individuals and their environments take on increased significance.

10.4 Agency through Empathy: The Psychological Debate

In this section, I will explore how agency is a relational development and explain how God is found in the midst of this development towards agency. The research in this section is drawn from attachment theory, attachment theory being a psychological model, which presupposes that people seek a connected and secure base of support from their earliest years from primary caregivers. It was developed from psychoanalyst John Bowlby's (1969/1982, 1973, 1980) work on loss among children, this is an important model for explaining the development of agency. Bowlby supposed that attachment experiences coalesced into a certain self-states based on experiences in early infancy. And attachment theory has recently been researched in relationship to neurobiology (Siegel 1999/2012). Bowlby revised psychoanalytic drive theory—the metaphor that people were driven through a hydraulic system of tension and release—towards a relational model that proposed that persons were first of all relationship seeking and that warmth and connection could be even more valuable than food or shelter. Care, rather than food or sex, was now seen as the fundamental need.

The first through the third years of life is especially important in attachment terms. Bowlby theorized that attachment coalesced into certain "self-states" based

on experiences during this period. Self-states are different from feelings or emotions because they come to be organizing principles that seem to speak to "reality." He labeled distinct self-states as "secure, avoidant, ambivalent, and disorganized." Persons whose caregivers were withdrawn may have experienced "avoidant" attachments, persons whose caregivers were threatening could experience "ambivalent" attachment, and so on. This was tested by several experiments that measured an infant's emotional reactions to a caregiver's leave-taking. For strict attachment theorists, these models of attachment remained fairly stable across the life span. Attachment theory shows the importance of empathy in early life.

A close look at attachment research indicates that one's sense of agency develops early in life through relationships with other persons (Fonagy and Target 1997). This is important because understanding the empathic and interpersonal roots of agency helps us foster agency in personal relationships and across society. First, I claim that the brain's mirror neuron system—housed in the secondary motor cortex and responsible for reflecting the actions of others within the brain—prepares persons to interpret intentions in others and that this plays an important role in the development of agency (Klein and Thorne 2006). This research indicates, at the level of scientific analysis, what Suchocki suggested was theologically sound—that people are profoundly shaped by their relationships. Second, I maintain that agency develops indirectly. One sees one's actions mirrored in another's behavior and then agency is attributed to oneself.

As noted in the introduction, the controversy in psychology about top-down versus bottom-up agency has proved to be a contentious issue, and my contribution shows how the capacity for top-down agency is actually the result of bottom-up processes and is thus fragile and malleable. Top-down agency approaches emphasize how a person is goal directed and able to act in accordance with certain values and intentions; bottom-up agency emphasizes how persons respond to stimuli in a role-determined fashion and underscores how agency is heavily determined by environmental factors (Prinz 2012). The distinctive argument of this section is that bottom-up stimuli are very significant, but that these bottom-up stimuli are personal rather than impersonal forces.

The stimuli that infants experience in relationships with primary caregivers—stimuli here being understood as the inherent drive towards relationship that exists within persons—provides the basis for later top-down agency. Top-down agency in a person is formed by a community or network of persons engaging in top-down agency with a person who is not able to engage in top-down agency on their own behalf (e.g. an infant not yet able to speak), highlighting the shared nature of personal agency. Agency is thus, from inception, a shared phenomenon built on relationality. Top-down agency has a developmental history in relationships of care.

Developments in neuroscience have further confirmed the interpersonal nature of agency, as described by attachment theorists. While attachment describes the capacity for agency as a form of "joint attention" that develops from particular relationships, neuroscientific findings describe how this capacity for agency is rooted in the brain's neuronal wiring (Siegel 1999/2012, p. 322). This helps explain the very early phenomenon of "mind reading" in which infants and young children

seem capable of not only seeing a person's actions but also intuiting the intentions behind a person's actions. The discovery of the mirror neuron system by a team of Italian scientists in the 1990s heralded a sea change in our consideration of what was happening in cognitive processing. Recent research in the cognitive sciences on empathy has built on these findings (Stueber 2006, p. 132). Specifically, mirror neurons contribute strong evidence for how our lives are linked with the lives of others through the inherently empathic capacities of our brains.

Mirror neurons are those parts of the brain that are able to witness another's facial expression and instinctively respond similarly. Capacities rooted in the mirror neuron system seem to take us a long way towards explaining how infants imitate facial expressions. An infant seems to mirror another's face without necessarily knowing of their own (Stump 2010).

In other words, the mirror neurons in an infant's mind fire almost automatically as they witness another's behavior; they do not seem to fire by accessing similar memories and attributing intention to another through comparing their inner states with the inner states of another. Mirror neuronal capacity seems more basic than cognitive processing. Mirror neurons thus appear to be nearly reflexive responses. "John grasps Mary's action because even as it is happening before his eyes, it is also happening, in effect, inside his head … mirror neurons permit an observed act to be directly observed by experiencing it" (Stump 2010, p. 69). Experience seems to *happen* in a child's mind before it becomes the object of observation, so that early shared experience could be described as preconceptual. Although it is difficult to describe how this interpersonal, second-person knowing is created through mirror neurons firing, it seems that infants reflect in some significant ways the minds of others. Perhaps the best that can be said at this point is that their minds respond to the minds of others by using the mirror neurons to create similar states in their own minds.

The discovery of mirror neurons has contributed significantly to our understanding of the capacity for social behavior and how this capacity is provided for in what seems to be the brain's structure. Specifically, recent research shows how this area of the brain functions in understanding the intentions and emotions of others. Research on neurological damage in which mirror neurons have been obliterated indicate that a person's capacity to understand the significance of another's actions is severely compromised (Preston and de Waal 2002). These persons can see others acting but cannot make accurate inferences about their intentions from observing their actions; a crucial "mind-reading" capacity seems to be missing.

The discovery of mirror neurons seems to indicate that the conceptual capacity for interpretation is built upon early aspects of the brain's functioning as it attributes intentions or emotions to others, intuiting the significance of their actions preconceptually. Mirror neurons explain why attachment is a more basic biological process than language acquisition and is more central to survival. This also accounts for why top-down agency is the result of bottom-up processes of interaction and imitation.

Mirror neuronal capacities challenge older cognitive and conceptual models based on meaning making. A meaning-making approach divides the self from the other and insists that interpretation is crucial. Mirror neuron research shows that

such self/other understandings are based on earlier self–other experience. Before meaning making, there is shared experience. When two persons are attending to the same thing, they are linked at the neuronal level even though they may not be consciously aware of it. This is not adequately explained by either first-person pronouns (the language of the "I") or third-person pronouns (the knowledge "that" something is the case). Suchocki's description of the web of human relationships indicates the profoundly connected interpersonal reality that is not completely captured by typical dualistic language for "subject" and "object," for self-experience and other experience.

This research indicates that "self" experience is based on "other" experience; we come at a self indirectly through the repetitive engagement with others who reflect ourselves back to us. As we have seen, recent discoveries in brain science suggest that persons engage in relationships at a preverbal level and that this capacity to match what is happening in another person's mind is an early precursor to the development of the self. Witnessing the actions and responses of others before having a knowledge of oneself as a separate entity, the preverbal experience of the mirror neuron system within individuals provides the substrate for the later development of what might be called a self. This matters for the debates about agency because it indicates that later top-down functioning is built on bottom-up phenomena in early life: through neuronal registering of the feelings, emotions, and self-states of others, one builds self-understanding.

Therefore, a practical and pastoral concern arises from this reflection: It matters a great deal whether we believe that others have agency, because, in believing this, we communicate this reality through our relationship to them. If persons first experience the faces of others and interpret their actions almost automatically based on the firing of their mirror neurons, they are inextricably bound into networks of shared experience. These networks of shared experience first offer certain kinds of self-states to a person and then later are used to construct meaning, convey values, and express sentiments. Expressions that affirm agency have the capacity to create the self-perception of agency within a person.

Recent psychological research links agency, top-down processing, and the development of goal-directed action with the experience of mirroring, suggesting that, from its inception, agency is inherently social. A joint achievement, agency develops through a process of mirroring and thus occurs indirectly rather than being a property of an individual. This suggests that what might have been previously considered as a trait or possession—the capacity to act with agency—is now seen as having a developmental basis. Furthermore, this developmental achievement is rooted in the brain's capacity for relational knowing seen in the mirror neuron system.

Although the primary developments in this line of thought have occurred in attachment theory, influenced by neurobiology, its results are starting to impact cognitive psychology as well. For example, the cognitive psychologist Wolfgang Prinz notes how persons observe each other's behaviors and notice not only the actions but implicitly come to conclusions about the reasons for and intentions behind them (Prinz 2012, p. 108).

Prinz ponders the origin of goals, which he calls "representations" in a person's goal-directed actions (Prinz 2012, p. 207). Asking a series of probing questions, he wonders how it is possible to have a goal in a horizon beyond oneself (Prinz 2012, p. 126). How do we "form explicit representations of events that are independent from the configuration of currently given circumstances?" (Prinz 2012, p. 126) In giving answers to these questions, he reserves a strong place for stimulus-based responses in shaping the world but does so without being reductionist. Although he writes in functionalist metaphors, he gives a surprisingly personal account of the development of "representations" from the stimulus of their lives. Bottom-up stimuli are the detour through which top-down agency is developed.

Prinz argues that agency begins with "perception" and develops in relationships (Prinz 2012, p. 136):

> Individuals indeed develop and implement top-down control from outside to inside, going from interpretation of … action [outside of themselves] (perception) to selection of their own action (production). First, they attribute agency and agentive control to others whom they see acting, and then, upon seeing others mirroring their own actions, they appropriate agency for themselves, attributing agentive control to their own actions. Eventually, these intuitions of agency lay the foundations for mechanisms of agentive (top-down) control of their actions, thus turning perception of agency into the production of goal-directed action. (p. 136)

Prinz describes a highly subjective process through which a person first experiences "intuitions" of agency by seeing his or her actions mirrored by others, and these later develop into the goals that shape goal-directed actions. The capacity for action is thus a fragile, contested, and interpersonally complex event. In this sense, agency is the result of stimuli, but stimuli that are inherently relational, the same that seek connection and warmth in patterns of attachment. In his analysis of the development of goal-directed behavior, he highlights how agency develops through the perception of others.

The mirror neuron system allows for second-person knowledge. Prinz shows how this system of witnessing another person's experience as if it were one's own is at the root of the development of personality and the capacity for developing goals that can lead to goal-directed action. "Individuals become willing agents by appropriating for themselves what they have first attributed to others" (Prinz 2012, p. 137).

Agency, goal development, and goal-directed action come from witnessing others witnessing oneself. Indirectly, the person comes to believe they have the power to act as they see their actions affecting the world. If their actions do not affect the world, they do not attribute agency to themselves. As neuroscience and attachment theory each indicate, the split-second processing of both action and intention is deceptively simple since it arises from a complex process of self-development.

When people are capable of top-down agency, or acting with intention, it reflects an interpersonal reality. Mirror neurons fire almost automatically in response to others and this sets up the conditions through which persons witness cues of safety and reliability in the faces of others from a young age. This perception includes the relational experience of agency: Another person confirms that they see us acting with

intention and according to certain values or desires. When this happens, we can begin to experience what Prinz calls "representations," or goals (Prinz 2012, p. 207).

I claim that the mirror neuron system helps explain how this works at a preverbal level: The infant witnesses cues in the faces of persons close to them and interprets the clues to develop a sense of safety. These cues include reflecting an infant's behavior, conveying to the infant a sense of self-reality and developing the capacity for action that later develops into the perception of agency. This early experience is thus a relational experience; people come to self-definition through connections with others. Over time, this consistent set of responses develops into a form of patterning that seems to appear fairly stable and allows for a platform from which a relational sense of self—a person acting with values and intentions—is first experienced. Agency depends upon those who witness our lives and behaviors and mirror back to us that what they see us doing is meaningful (Stern 1985/2012).

In this section, we have explored how relationships are fundamental to what could be called the development of "self," especially as it regards the self-perception of agency. Indeed, it matters a great deal whether we believe other persons have agency because in holding that belief we communicate that to them through the shared experience that we have with them.

10.5 Conclusion

Persons are connected in multiple relationships from before birth, and it is primarily in these relationships that they develop a sense of agency. In the previous section, we have explored how what we consider a "self" is actually born out of relationships of mirroring in which another reflects one's actions and that the mirror neuron system allows this action to be reflexive—to be interpreted by the self. From this perspective, agency is seen to be inherently relational. Although this occurs foundationally in infancy, it also occurs across the life span as relationships of attachment can change through the course of the life span, with persons coming to have increasing agency. Suchocki highlighted how there are systems of bonding that nourish us before and after we are born, long before we can use the name "I" to describe ourselves, much less attribute agency to ourselves.

To put a fine point on it, agency is a capacity that develops with the empathy of others through actions of joint attention in which persons share experiences. Without these mirrors for experience, people do not perceive themselves as having agency as easily. Those who have been systematically denied agency frequently still manage to attempt to achieve some sense of agency—reaching "horizontal self-transcendence," but even this capacity to search for agency is built upon a network of relationships in one's life that constituted a source of self-imaging.

God is proximate to these networks of relationships, these communities of well-being, through which persons discover themselves as having agency and thus are able to develop meaningful goals and work towards them. The human processes of joint attention, empathy, and attachment could be seen as "natural" processes

occurring in relationships of care, given my thesis about the linkage points between persons being where God is present, these processes are also the distinctive location of human agency towards connected communities of well-being.

The consequences of my argument that we are neurologically "wired" for shared experience but that that which we see can only shape us (Ellison 2013). There are several important consequences that could be drawn from this, and further work that needs to be done on the effects of "horizontal self-transcendence" within a relational psychological/theological understanding of agency in regard to particular social problems.

In this chapter, I have explained how agency has been understood in the Christian tradition and staked a claim in the feminist process theological understanding of agency, namely that people are capable of "horizontal self-transcendence" when they use their agency to foster interconnected communities of well-being in the world (Suchocki 1994, p. 35). Using attachment theory, I have shown how important the second aspect of horizontal self-transcendence, namely *empathy*, is to the development of agency. Implicitly, I have shown how this view of human freedom is rooted in a rich theological tradition in which the agency of the self is meant to be in service to others and to God. Because people are born into webs of relationships that systematically deny the agency of others, this should be understood not only as a violation of others but also as a violation of God as God is witnessed in the linkage points between persons.

Perhaps the most challenging claim I have made in the chapter is that agency is inherently relational. Agency is not a personal possession, an essentialized attribute of character that can be deployed at will. Instead, agency is a shared phenomenon so that persons are responsible to foster agency not only for themselves but also for others who have been systematically denied agency. Using process theology as a theological framework, I have argued that, although these relationships are important, social and interpersonal relationships also reflect upon God. Instead, I maintained that these relationships amount to a violation of God since each act of interpersonal violence, and each instance of denial of agency, is felt by God as an infringement against the well-being of a person. God is the one who influences the world towards justice and well-being for all, using God's agency to attempt to change persons who are capable of responding to God's influence. Since God is seen as that force existing in and through relationships to foster agency and well-being at the intersections of persons, society, and nature, fostering agency is a shared task that is profoundly theological, thus deserving to be at the center of pastoral theological reflection in the twenty-first century.

References

Ashley, J. M., & Metz, J. B. (2007). *Faith in history and society: Toward a practical fundamental theology*. New York: Crossroad Pub. Co.

Beste, J. E. (2007). *God and the victim: Traumatic intrusions on grace and freedom*. Oxford: Oxford University Press.

Bowlby, J. (1969/1982). *Attachment*. New York: Basic Books.
Bowlby, J. (1973). *Separation: Anxiety and anger*. New York: Basic Books.
Bowlby, J. (1980). *Loss: Sadness and depression*. New York: Basic Books.
Browning Helsel, P. (2009). Gender Roles in the encyclopedia of psychology and religion (p. 344–346). In D. A. Leeming, K. Madden, & S. Marlan (Ed.), New York: Springer.
Chopp, R. S. (2007). *The raxis of suffering: An interpretation of liberation and political theologies*. Eugene: Wipf and Stock.
Ellison II, G. C. (2013). *Cut dead, but still alive: Caring for African-American young men*. Nashville: Abingdon.
Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role inself-organization. *Developmental Psychology, 9,* 679–700.
Gutiérrez, G. (1973). *A theology of liberation: History, politics, and salvation*. Maryknoll: Orbis Books.
Hartshorne, C. (1948). *The divine relativity: A social conception of god*. New Haven: Yale University Press.
Hartshorne, C. (1984). *Omnipotence and other theological mistakes*. Albany: State University of New York.
Hoeft, J. M. (2009). *Agency, culture, and human personhood: Pastoral theology and intimate partner violence*. Eugene: Pickwick Publications.
Hollenbach, D. (2002). *The common good and christian ethics*. Cambridge: Cambridge University Press.
Iber, S. T. (2010). *Principle of subsidiarity in catholic social thought: Implications for social justice and civil society*. New York: Peter Lang.
Johnson, E. A. (1992). *She who is: The mystery of god in feminist theological discourse*. New York: Crossroad.
John Paul II, *Laborem Excersens*. 14 Sept. 1981, 3 Oct. 2013, Web. http://www.vatican.va/holy_father/john_paul_ii/encyclicals/documents/hf_jp-ii_enc_14091981_laborem-exercens_en.html
Jones, S. (2000). *Feminist theory and Christian theology: Cartographies of grace*. Minneapolis: Fortress Press.
Klein, S. B., & Thorne, B. M. (2006). *Biological psychology*. New York: Worth.
Luther, M. (1970). *Three treatises*. Minneapolis: Fortress.
Niebuhr, R. (1941/2004). *The nature and destiny of man: A christian interpretation*. Louisville: Westminster John Knox.
Prinz, W. (2012). *Open minds: The social making of agency and intentionality*. Cambridge: MIT Press.
Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25,* 1–72.
Phillips, D. Z. (1965–2014). The Concept of Prayer. London: Routledge.
Rigby, C. (2001). All God, and us: Double agency and reconciliation in Psalms 22 and 51. In S. Breck Reid (Ed.), *Psalms and practice: Worship, virtue, and authority* (pp. 202–219). Collegeville: Liturgical.
Siegel, D. (1999/2012). *The developing mind: How relationships and the brain interact to shape who we are*. New York: Guilford.
Stern, D. (1985/2012). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. London: Karnac.
Stueber, K. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences*. Cambridge: MIT Press.
Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford: Oxford University Press.
Suchocki, M. H. (1994). *The fall to violence: Original sin in relational theology*. New York: Continuum.
Swinton, J. (2012). *Dementia: Living in the memory of God*. Grand Rapids: Wm. Eerdmans.

Chapter 11
Agency and Self-Experience: The Religious Function of the Psyche

Robert Childs

While certainly no theologian, my commentary on Browning Helsel's (Chap. 10, in this volume) paper on the *Relational Basis of Agency* will be predicated on the overlap between psychology and religion as an alternative viewpoint on the psychological development and utility of agency. Because Browning Helsel's perspective on agency is based on process theology with its focus on shared agency fostered in the network of relationships with others and with God, juxtaposed to a sense of agency that he refers to as "top-down" and goal-focused agency, I will attempt to offer a third view that I believe addresses some of the psychological functions that agency serves. Finally, I hope to draw some parallels between process theology and Jungian depth psychology that may offer a new way of holding the tension between divergent points of view.

Let me start with one of the central questions in this finely argued and well-developed paper: What are the roots or source of human agency and ultimately what is the function of agency itself? From Dr. Browning Helsel's perspective, referencing Ellison (2013), "Agency—the feeling that one can do something to influence the world—arises from experiences of being seen and of witnessing one's own actions mirrored in another's response" (Chap. 10, in this volume, p. 3). From the process theological viewpoint, God is integral to relation so that the exertion of human agency impacts God. In other words, there is a consistent empathy (focus on the other) in which God is revealed and this becomes the ground out of which agency is formed. People participate with God when they use their agency for the well-being of others. Therefore, agency is created through shared interpersonal experience rather than an individual achievement. From this perspective, agency does not imply autonomy but the effect of a network of relations.

Browning Helsel's emphasis here is on the interpersonal formation of agency and he references attachment theorists, as well as several of the postmodern intersubjective psychological theorists, such as Fonagy and Target (1997) and Stern (1985), to scaffold his argument that there is an overlap between a relational

R. Childs (✉)
Massachusetts School of Professional Psychology, Newton, MA, USA
e-mail: rchilds618@comcast.net

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_11

psychology sense of agency and its evolution into a theological pathway to God. He notes the agency that a person exerts does not occur outside of social constraints (or embodiment) but occurs within the web of relationships in which he or she finds themselves. From a psychological perspective, Browning Helsel is placing agency within the social constructivist camp of self, agency being developed out of interpersonal experience which he extends through his use of the mirror neuron system research (Prinz 2012).

There is, of course, a long-running discussion through the history of psychology about the development of self and agency. In attempting to move beyond Freud's notion of primary narcissism (the infant focusing his initial emotional interest entirely on himself) and Mahler's sense of the symbiotic or undifferentiated experience of the infant, Stern (1985) rejected these assumptions and suggested that they were secondary to and dependent upon an already existing sense of self and others. This was extended by Sander (1983) and other infant researchers who point out that the traditional view of the infant starting from the simple and building to the complex has been disproven; we now see the infant as an extremely complex living organism already possessing integrating mechanisms. For Stern (1985) and others in the Boston Process Change Study Group, the representation of self is built upon interaction. The self is built upon the memory traces of interpersonal events and, in order to be encoded and internalized, each experience must have an affective component. Here, the emphasis is on the external cueing function of the self and other relation through interaction, which results in mutual regulation.

It is understandable how this dovetails nicely with Browning Helsel's (2015) argument, especially in terms of the emphasis on the role of cognition and interpersonal relationships as well as the mirroring component of shared states in the development of self and agency. I would offer that what is missing in this view of the interpersonal development of agency and self are the internal cueing states of both the infant and the individual adult. For me, this postulate is related to the unfolding of what Slavin (1997), Gentile (1998), and Ogden (1994) have called a deep structure predisposition (Bollas (1989) referred to the same idea as a personal idiom) in a relationship to a specific transformational object. Simply stated, from my viewpoint, the self and agency are not just socially constructed.

This a *priori* or deep structure foundation can be thought to be a genetically biased set of dispositions that exist before object relating. Ogden (1994) describes this deep structure as "biologically organized templates that serve to organize the immense quantity of experiential data with which the infant or child is flooded." While this deep structure is never directly experienced, these biologically encoded, psychological predispositions are seen as awaiting creation through interaction. The self is not only created intersubjectively but is also created in the unfolding of what preexists in "conversation" with the other. So I would be in agreement with Stern and ultimately Browning Helsel that the infant is not passive, nor solipsistic. My emphasis, however, is that there is a self that is there already but which needs to be co-shaped through interaction. I will return to this later as I look at the role that archetypes and self-experience play in the development of agency.

For now, I would like to extend this perspective to explore a different viewpoint on the nature of agency. In framing my perspective, I would like to draw on the work of Schachtel and Winnicott as well as a larger humanistic understanding of agency. I would like to start by saying I have great sympathy for the perspective that Browning Helsel offers, as a top-down approach to agency that is goal oriented feels egoic and self-focused. At the same time, I am reluctant to focus only on shared agency. This will become clear as I attempt to contextualize an alternative perspective on agency grounded in humanistic theory and Jungian psychology.

In several papers throughout his career, Winnicott (1950, 1964, 1971) describes a process that locates agency and aggression as a necessary part of development. His central premise is that aggression is prior to the integration of the personality and is almost synonymous with activity. Winnicott is proposing here that there is a positive and creative value in aggression and destructiveness in a human need to develop a process of integration in the service of building a self. Another way of saying this is that the roots of agency lie in the ability of the child to say no; to be a separate and distinct individual while at the same time being a part of the web of connections. A key part of development is the necessity of separating out subjective experiences so we can then see ourselves. Even if the goal is shared agency or a unifying experience with God, separation is a necessary part of development. Perhaps this is why Adam and Eve were thrown out of the Garden of Eden after eating from the tree of knowledge; perhaps this was their original sin. As Campbell (1988) points out, this was an act of disobedience ultimately seen as the beginning of human initiative or agency.

Adding in this neglected aspect of the function of aggression can shift the way we see Browning Helsel's use of Suchocki's (1994) perspective that a kind of interpersonal violence surrounds persons before their birth, and thus could be considered a form of "original sin" in that it is a form of the violations of others that precedes personal responsibility. In this context, agency becomes an important act of resistance against the pressure for social conformity or identity expectations (Helsel 2014, p. 16).

Schachtel (1959) notes that the word "aggression" stems from the Latin aggredi, meaning to go towards or approach. We have lost this aspect of the meaning of aggression. Aggression, activity, and agency now become grounded in the need to shape experience on behalf of one's own self-interest in addition to its use as hostile attack. This is not to deny the reality of oppression or the denial of those who have not had the opportunity to have a self; rather it is to show a new way of seeing how within this inalienable right of selfhood for all that there is a way to show how the individual's active and creative shaping of experience requires the "destruction" and "reconstruction" of certain aspects of life. This process is captured nicely in a story told by religious historian Mircea Eliade (1985) who, while observing a painter in India, once remarked, "first you have to create the world, then you destroy it, then you create it again, then you destroy it again and so on until you end up with a painting."

I also have difficulty with a top-down approach that is focused only on the capacity of agency in the service of separate and individual goals. I would suggest

instead that individual agency is hard won but, in fact, is not the goal. I would also suggest that this function may be extended much in the way Epstein (1988) talks about the need for a synthetic ego, based on Janet's focus on the synthetic aspect of healthy mental functioning. From Epstein's point of view, the ego can be seen as a matrix of structures, not as a single entity, with both functional and representational aspects. Psychotherapists know this experience of representation as the capacity to have their attention "evenly hovering" in the service of observing multiple layers of experience. The essential element of this middle path between individual agency and shared agency is the development of a capacity to see beyond what is literal and concrete; to see into what the religious historian Corbin (1971) called the mundus imaginalis, the reality of the imagination. Here, we could see Browning Helsel's effort to extend agency in the service of connecting to a wider web and ultimately to God as parallel to the Jungian effort to place the psyche in the service of connecting the individual ego to the larger world. As Jung pointed out, our fantasy is that the psyche is in us, but in reality we are in psyche. At the same time, Jung would be the first to point out that it is essential that the ego remain anchored in consciousness and not be assimilated by the psyche.

Jung's concept of a transpersonal Self is different from theories of a personal egoic self, which is usually seen as derived by a process of internalized accumulation of experience. As Jung himself said in the Red Book (2009): "In as much as the I is only the center of my field of consciousness, it is not identical with the totality of my psyche, being merely a complex among other complexes. Hence I discriminate between the I and the Self, since the I is only the subject of my consciousness, while the Self is the subject of my totality, hence it also includes the unconscious psyche." This expanded notion of Self attributes an intelligence to the psyche (Self) which communicates to the I/ego through the medium of images. For Jung, the Self constitutes the most immediate experience of the divine. As the Jungian Corbett (1996) notes, "The Self is the slow gradual realization of the divine cosmic center in the unconscious psyche of the individual."

To help create a bigger context for moving beyond the process theological framework that agency is based on relational experience, I am brought to a place of great resonance with Dr. Browning Helsel's paper, which is in the overlap between process theology and Jungian psychology. The place to start with this is the conference held at Claremont College in 1983 (Griffin 1989), which explored the relationship between process theology and archetypal psychology. Among the parallel themes that emerged from the conference were the emphasis on creative activity, rather than passive matter, and an evolutionary becoming rather than a changeless enduring perhaps best expressed through the concept of dipolar theism: "the idea that God has both a changing aspect (God's existence as a living god) and an unchanging aspect (God's eternal essence)." Both archetypal psychology and process theology want to return soul and divinity to the world. Jung's emphasis was on the soul's knowledge and Whitehead's emphasis was on the world's self-knowledge.

In his book, *Aion* Jung (1951) says that "Christ exemplifies the archetype of the Self" and is therefore worthy of imitation. This deep parallel with process theology

is exemplified with Whitehead's notion of God as primordial, which is the locus of eternal objects. As Tarnas (2006) points out "archetypes can be understood and described in Whiteheadian terms as eternal objects and pure potentialities whose ingression informs the unfolding process of reality." In other words, God may be thought of as the archetype of archetypes, which makes contact with Whitehead's notion of God as primordial … which is the locus of eternal objects (Slusser 1989). Maxwell (2011) suggested that Whitehead's cosmology can provide philosophical grounding and justification for Jung's psychology of the archetypes.

One final parallel to mention regarding the overlap between process theology and Jungian psychology is the idea that matter and energy have both exteriority and interiority and that these two realms are intimately connected (Griffin 1989). Here, the panexperientialism of process theology can been seen reflected in Jung and Hillman's (1975) characterization of psyche and the unconscious. Browning Helsel's words echo this idea when he says "… influenced by Alfred North Whitehead: God should be understood as that force within creation that influences all creating towards becoming." The synonymous implications with psyche are obvious.

The shift in perspective I offer is to widen the external relational focus of process theology to include or extend the web of relationship to our inner world. This gives us an opportunity to move beyond psychoanalytic reductionism; our internal world is more than just drives, instincts, and egoic goals. It also allows us to shift the notion of shared agency from the exterior to the interior, as one begins to give the Self agency in the unfolding of becoming. From a Jungian point of view, this shift is captured in a more expansive relationship between the egoic realm and psyche. "Sensing the Self as something irrational, as an indefinable existent, to which the ego is neither opposed nor subjected but merely attached and about which it revolves very much as the earth revolves around the sun, thus we come to the goal of individuation. The individuated ego senses itself as the object of the unknown and supra-ordinate subject" (Jung 1951). The psyche is now the initiator or agent of the individual's life, much as Browning Helsel is suggesting Christ can bring meaning to the individual through shared agency. The psyche can provide the internal cueing states that provide a synthetic function of re-establishing a new equilibrium for the individual.

Jung long appreciated the irreplaceable role that the urge towards wholeness and self-actualization plays in development. Jung himself placed at the center of his psychological theory his sense that humans were teleological beings who gained access to the religious dimensions of experience through the symbolic, imagistic, and mythological expressions of their psyches. Jung came to call this process of the linking and influence of the unconscious psyche on individual consciousness "individuation" and once even said that it "can be seen as the primitive Christian idea of the Kingdom of Heaven which is within you" (Jung 1966). To me, this parallels Browning Helsel's framing of shared agency as beyond the individual. As Jung (1966) once said, individuation "… brings to birth a consciousness of human community precisely because it makes us aware of the unconscious, which unites and is common to all mankind. Individuation is an at-one-ment with oneself and at the same time with humanity, since oneself is a part of humanity."

If Christ is seen as an archetype of the Self, a totality of the divine, then as Chapman notes, the life of Christ is identical in us, from the psychological point of view, with the unconscious tendency towards individuation. Chapman (1997) goes on to say that Jung believed this supreme archetype (of Christ) is latent in each human being, waiting to be manifested consciously through a lifelong struggle to balance the various forces of the psyche. That process cannot merely be taught, but involves taking personal responsibility for the life forces within oneself and holding together with their opposition. Eventually, "it can happen only when you withdraw your projections from an outward historical or metaphysical Christ and thus wake up Christ within … thereby we become truly human as well as participants in the divine life, for responsible living and fulfilling of the divine will in us be our form of worship and commerce with God" (Chapman 1997). This reflects the realization that the Self lives in you and not in an external figure separated and different from yourself. Said another way, God lives in everybody and with an internalized sense of agency that comes from this relationship perhaps we have ended up in a place close to Browning Helsel but with the added perspective on the internal aspects of agency.

A place of potential divergence between process theology and Jungian thought is the question of how one develops a relationship with the living God. To understand the nature of Logos in process theology, I think we have to step back and look at a few ideas central to Alfred North Whitehead's process philosophy. Whitehead (1929/1978) understood symbolism as an interpretive judgment, which required direct experience. He coined the term "prehension" (from the Latin prehensio, to seize) by which he meant that the perceiver actually incorporates aspects of the perceived thing into himself. His use of the term "symbolic reference" was Whitehead's way of not trusting symbolic interpretation. He suggested that beyond causal efficacy (or physical prehension), a second mode existed which he referred to as presentational immediacy, which is often thought of as pure sense perception. Prehension is a nonsensory mode of perception which may or may not be conscious. From Whitehead's point of view, people can perceive automatically using "causal relations between entities," without deeper reflection.

For Whitehead, nothing exists without being in relation with everything else, including God, which I imagine is why Browning Helsel places his emphasis on the web of relationships. However, as Cobb and Griffin (1976, p. 100) note:

> Creative transformation is the essence of growth and growth is the essence of life. Growth is not achieved by merely adding together elements in the given world in different combinations. It requires the transformation of those elements through the introduction of novelty … the source of novelty is the Logos, whose incarnation is Christ. Where Christ is effectively present, there is creative transformation.

Key for process theology is that God is seen as creative love, which is identified with the primordial nature or Logos of God. The question becomes, however, how to access the primordial nature of God. Here, Jung was clear that the best way to access primordial experience is through symbols, images, and mythology. He felt "the term religion designates the attitude peculiar to all consciousness which has been changed by an experience of the numinosum" (Jung 1938). Jung's answer to Whitehead's concern for symbolic reference was closer to his understanding of

prehension. He would think in terms of the religious nature of the psyche in terms of its Latin etymology, *religare,* to bind. His efforts were to name in images rather than concepts, which brought the individual directly in touch with primordial nature. Jung's psychology could best be characterized as the love of images.

As Jung (1963) says in his autobiography, "it is not that 'God' is a myth, but that myth is the revelation of a divine life in man. It is not we who invent myth, rather it speaks to us as a word of God." For Jung, the archetypal, primordial images of myth are the language of the unconscious or psychic processes and thus point us towards symbolic knowledge and ultimately our divine nature. "No intellectual expressions comes anywhere near the richness and expressiveness of mythological imagery" (Jung 1944). So it is through developing a relationship with his psyche that the individual can get in touch with creative transformation. This way we can incorporate or prehend the creative expression of the psyche into ourselves. The search for wholeness and patterns in the Logos, what Whitehead would refer to as the "absolute wealth of potentiality," has resonance with what Jung saw as the power of archetypes and symbols (Kling 2011). As Jung (1975) says, we become truly human as well as participants in the divine life, for "responsible living and fulfilling of the divine will in us will be our form of worship of and commerce with God." Through discovering the expression and agency of our psyches, we can then use our new perspectives for the well-being of both ourselves and others.

References

Bollas, C. (1989). *Forces of destiny: Psychoanalysis and human idiom*. London: Free Association.

Browning Helsel, Philip (2015). The Relational Basis of Agency: An Integrated Psychological/Theological Approach. *Annals of Theoretical Psychology* (Vol. 12, pp.).

Campbell, J. (1988). *The power of myth (with Bill Moyers)*. New York: Doubleday.

Chapman, G. (1997). Jung and christology. *Journal of Psychology and Theology, 25*(4), 414–426.

Cobb, J. B., & Griffin, D. R. (1976). *Process theology: An introductory exposition*. Louisville: Westminster John Knox.

Corbett, L. (1996). *The transpersonal self in the religious function of the psyche* (pp. 39–56). London: Routledge.

Corbin, H. (1971). Mundus imaginal, or the imaginary and the imaginal. In *En Islam Iranien: Aspects spirituels et philospophiques* (onrne IV, livre 7). Paris: Gallimard.

Eliade, M. (1985). *Symbolism, the sacred and the arts*. New York: Crossroad.

Ellison, G. C. II (2013). *Cut dead, but still alive: Caring for African-American young men*. Nashville: Abingdon.

Epstein, M. (1988). The deconstruction of the self, ego and "egolessness" in Buddhist insight meditation. In *Psychotherapy without a self, a Buddhist perspective*. New Haven: Yale University Press.

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Developmental Psychology, 9,* 679–700.

Gentile, J. (1998). Listening for deep structure: Between the a priori and the intersubjective. *Contemporary Psychoanalysis, 34*(1), 67–89.

Griffin, D. R. (1989). *Archetypal process: Self and divine in Whitehead, Jung and Hillman*. Evanston: Northwestern University Press.

Hillman, J. (1975). *Re-visioning psychology*. New York: Harper and Row.

Jung, C. G. (1938). *Psychology and religion*. New Haven: Yale University Press.
Jung, C. G. (1944). *Psychology and alchemy*. Princeton: Princeton University Press.
Jung, C. G. (1951). *Aion: Researches into the phenomenology of the self* (trans. R. F. C. Hull). New York: Pantheon.
Jung, C. G. (1963). *Memories, dreams, reflections*. New York: Vintage.
Jung, C. G. (1966). *Two essays on analytical psychology*. Princeton: Princeton University Press.
Jung, C. G. (1975). Letters. In G. Adler & A. Jaffe (Eds.), *On behalf of the mystical fool: Jung on the religious situation*. Princeton: Princeton University Press.
Jung, C. G. (2009). *The red book*. New York: W. W. Norton.
Kling, S. D. (2011). Resonances between process and Jungian thoughts especially in the christology of Cobb and Griffin. A talk given at the 2011 student symposium on science and spirituality at the Lutheran school of theology at Chicago. 1–13.
Maxwell, G. (2011). Archetype and eternal object: Jung, whitehead and the return of formal causation. *Archai: The Journal of Archetypal Cosmology, 3,* 51–71.
Ogden, T. (1994). *Subjects of analysis*. Northvale: Jason Aronson.
Prinz, W. (2012). *Open minds: The social making of agency and intentionality*. Cambridge: MIT Press.
Sander, L. (1983). To begin with—Reflections on ontogeny. In J. Lichtenberg & S. Kaplan (Eds.), *Reflections on self psychology*. Hillsdale: Analytic.
Schachtel, E. (1959). *Metamorphosis*. New York: De Capo.
Slavin, M. (1997). Personal Communication.
Slusser, G.H. (1989). Jung and whitehead on self and divine: The necessity for symbol and myth. In D. Griffin (Ed.), *Archetypal process: Self and divine in whitehead, Jung and Hillman*. Evanston: Northwestern University Press.
Stern, D. (1985). *The interpersonal world of the infant*. New York: Basic.
Suchocki, M. H. (1994). *The fall to violence: Original sin in relational theology*. New York: Continuum.
Tarnas, R. (2006). *Cosmos and psyche: Intimations of a new world view*. New York: Viking.
Whitehead, A. (1929/1978). *Process and reality*. New York: The Free Press.
Winnicott, D. W. (1950). *Aggression in relation to emotional development in through paediatrics to psychoanalysis* (pp. 204–218). London: Tavistock.
Winnicott, D. W. (1964). *Aggression and its roots in deprivation and delinquency* (pp. 73–86). London: Tavistock.
Winnicott, D. W. (1971). *The use of an object and relating through identifications and creativity and its origins in playing and reality* (pp. 65–94). London: Tavistock.

Chapter 12
Psychology's Purgatory: Situating the Theoretical Construction of Agency

Jaan Valsiner

Let us begin from the basic axiom—the *notion of* God (or any version of deities in any societies) *is not a theoretical concept* in scientific psychology. It can be a very important semiotic sign complex to organize human lives (and deaths) of millions of people worldwide, it may be a commonsense notion for which many people want, or agree, to die. Despite all of its social and psychological importance, that notion is not a concept with which psychological theories can operate. It is a notion that needs to be explained itself—from the perspective of psychological theories.

12.1 Religious Phenomena are Crucial for Scientific Psychology

All phenomena of religious kind—and there is a mindboggling multitude of those all around the world—are relevant phenomena for psychological theories to address. The ease of creating individual deities as helpers in psychotherapy process (Valsiner 1999) or Alice Lakwena's creative adjustment of the Christian imperative "you shall not kill!" to the ways of warfare of the Lord's Resistance Army (Behrend 1999) are extremes of the examples how the complex semiotic-mediating devices we indiscriminately call "religions" matter in the living and dying of real human beings. Both destruction and construction are accomplished through the help of religious systems ranging from crusades and iconoclasms to personal retribution efforts (Obeyesekere 1975). Children are brought up with a focus on expecting some religious figures to perform miracles (Josephs and Valsiner 1999; Watzlawik and Valsiner 2012).

When seen from the perspective of cultural psychology, all religious systems in the World emphasize one or another kind of *transcendentality persuasion* upon

J. Valsiner (✉)
Niels Bohr Professor of Cultural Psychology, Department of Communication and Psychology, Aalborg University, Aalborg, Denmark
e-mail: jvalsiner@gmail.com

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_12

the upcoming generations. Human beings are persuaded that *they themselves* need to have the will to believe in agency that is located beyond the borders of their immediate accessibility. When a young child is told a "miracle story"—about Jesus walking on the water—the immediate contrast with the child's own inability to do so is obvious. Yet, the generalization suggested from that contrast links the unable (the child) with the able (Jesus) through the suggestion "if *you believe enough* you could be like Jesus." The self-motivating intra-psychological goal directedness is set up—the child who accepts the semiotic trap may start on the path on believing more and more—trying to reach the status of "enough" so that she or he could replace the mechanical water-skiing by spiritually based walking on the water. The latter moment never comes—yet many other psychological changes happen in this process of trying to reach the horizon.

What is at stake here is the infinity of the agentive role of human beings—the imperative of "you should believe more" and "even more" and "even more"—and so to infinity (or a stop in that process—at death or conversion to atheism, yet another infinite belief system). Psychology needs to create a theoretical understanding of such infinite belief systems that reach a high level of extension in time—as shown by history of all religions—and generalizing abstraction. The religious domain is the best arena for study of processes of hyper-generalization of signs (Beckstead et al. 2009; Valsiner 2014). This process leads to psychological phenomena of complete "takeover" of the human minds (and hearts), and is thus the ultimate empirical research object for general psychology.

12.2 Going Beyond William James

It is James' *Varieties of Religious Experience* (James 1902) that is usually cited if a psychologist wants to present oneself as a sophisticated researcher in the arena of complexities of religions. The focus on *experience*—personal feeling-in into the world—was crucial for James. Interestingly, psychology of religion had, from the 1920s onwards, become a hostage to empirical comparisons of religious groups of various kinds, as to their average results on different psychological "measures." What had become lost in the process of turning psychology of religion into a field of "empirical research" was precisely the focus on *experience*—the issue that both of the contributors to this section of our book (Browning Helsel; Chap. 10, this volume; Childs; Chap. 11, this volume) emphasize. This restoration of the focus on experience needs to be situated in the context of the reemergence of psychology of religion—now in the wider framework of cultural psychology (Belzen 2010). The coverage of religious experiences is crucial for turning psychology into a comprehensive science where the war; between the natural sciences and *Geisteswissenschaften* can be broken. Not including religious experiences—together with other hyper-generalized values (Branco and Valsiner 2012)—would be an obstacle for psychology as science. The latter is still to be reached, since William James'

depiction of the state of the art (science) of psychology given in the year 1892 can still be true today:

> Psychology … is to-day hardly more than physics was before Galileo, what chemistry was before Lavoisier. It is a mass of phenomenal description, gossip, and myth, including, however, real material enough to justify one in the hope that with judgment and good will on the part of those interested, its study may be so organized even now so as to become worthy of the name of natural science at no very distant day. (James 1892, p. 146)

Of course, psychology has progressed over the past century—yet in a direction that, replacing "a mass of phenomenal description, gossip, and myth" in James' quote by "a mass of *p*-values, pseudo-empirical theories, and unsupported claims of public usefulness" the delayed development of the area into a natural science is still obvious. Psychology has largely imitated natural sciences, rather than become one itself. In the natural sciences, it is the nature of the object phenomena that is honored—even if studied from various angles of approach that diminish the full richness of the phenomena. In psychology, phenomena are often "measured out of existence" by attributing numbers to them, using large samples rather than in-depth study of individuals. The phenomena of religious experiences have been left out of consideration for decades. Such experiences are central to human beings—as *Homo sapiens* operates at the highest level of reflexive self-organization far beyond the basic brain processes that fascinate our contemporary neuroscientists. Scientific study of these higher psychological functions is the task for cultural psychology—an up-and-coming discipline.

References

Beckstead, Z., Cabell, K. R., & Valsiner, J. (2009). Generalizing through conditional analysis: Causality in the world of eternal becoming. *Humana Mente*, 11, 65–80. [http://www.humanamente.eu]

Behrend, H. (1999). *Alice Lakwena & the Holy Spirits*. Oxford: James Currey.

Belzen, J. (2010). *Towards a cultural psychology of religion*. New York: Springer.

Branco, A. U. & Valsiner, J. (Eds.) (2012). *Cultural psychology of values*. Charlotte: Information Age Publishers.

James, W. (1892). Plea for psychology as a 'natural science'. *Philosophical Review, 1*(2), 146–163.

James, W. (1902). *The varieties of religious experience*. London: Longmans, Green, and Co.

Josephs, I. E., & Valsiner, J. (1999). Meaning-making and miracles: The creative inconsistency of the mind. In K. H. Reich, F. K. Oser & W. G. Scarlett (Eds.), *Psychological studies on spiritual and religious development*. Vol. 2. Being human (pp. 101–114). Lengerich: W. Pabst.

Obeyesekere, G. (1975). Sorcery, premeditated murder and canalization of aggression in Sri Lanka. *Ethnology, 14,* 1–23.

Valsiner, J. (1999). I create you to control me: A glimpse into basic processes of semiotic mediation. *Human Development, 42,* 26–30.

Valsiner, J. (2014). *Invitation to cultural psychology*. London: Sage.

Watzlawik, M. & Valsiner. J. (2012). The making of magic: Cultural construction of the mundane supernatural. In J. Valsiner (Ed.) (2012). *The Oxford Handbook of Culture and Psychology* (pp. 783–795). New York: Oxford University Press.

Part IV
Socio-ecological perspectives on agency

Chapter 13
Agency enabled by the *Psyche*: Explorations using the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals

Jana Uher

Nothing is more fascinating than the phenomena that we experience in every waking moment of our lives—each individual for him- or herself. These phenomena sometimes appear to be like windows through which we can peer into the world. At the same time, they also appear to constitute a world of their own within us, one that is interwoven with the world that we conceive of as being external to us and in which we exist as part of its universe. The unique phenomena of the psyche and their workings, intangible and ephemeral, have occupied philosophers, scientists and many others over the past millennia and even further back in the natural history of humankind. Psychical phenomena influenced human developmental pathways in unprecedented ways, producing the abilities that first enabled scientific explorations (e.g. Baldwin 1906; Fahrenberg 2008a, 2008b, 2013; Freud 1915; Hegel 1807; Hirschberger 1980a, b; Kant 1798; Peirce 1931–1935; Schrödinger 1958; Tomasello 2014; Walach 2013; Wundt 1863, 1894; Vygotsky 1978).

Explorations of these unique phenomena encounter profound challenges unknown to other explorations. For how can individuals explore and understand an object of investigation from which they themselves are inseparable (Durkheim 1919)? How can humans explore the human mind when they are equipped with nothing but such a mind (Stent 1969; Uher 2015a)? Scientists cannot step outside of themselves and of their being as human individuals. But scientists can explicate the basic assumptions that they have made about their particular objects of research and about the fundamental notions by which knowledge about them can be gained (Aristotle 350 BCE; Fahrenberg 2013; Collingwood 1940; Walach 2013). This enables critical reflection—individually and in exchange with other scientists. The ways in which individuals develop and exchange ideas, experiences and knowledge are explored in this research.

J. Uher (✉)
London School of Economics and Political Science, Department of Psychology, London, UK
Comparative Differential and Personality Psychology, Free University Berlin, Germany
e-mail: mail@janauher.com

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_13

13.1 The Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals

This research aims to explore the workings of the psyche by applying the *Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals* (briefly referred to as *TPS-Paradigm*). This novel research paradigm explicitly considers the challenges and limitations that are inherent to explorations of individuals and their minds and it aims to meet and minimise these challenges and limitations by adopting transdisciplinary and philosophy-of-science perspectives (Uher 2013, 2015a, b, c). Hence, the intention of this research is not to comprehensively review previous lines of research but rather to complement the existing knowledge with new insights that can be gained from the transdisciplinary and philosophy-of-science perspectives that are still not well considered.

The article first introduces the nature and aim of the TPS-Paradigm and specifies some philosophical presuppositions that the paradigm makes about individuals as objects of research. It then elaborates metatheoretical properties that the TPS-paradigm conceives for the phenomena explored in individuals, putting the main focus on the phenomena of the psyche. On the basis of these elaborations, the article explores the unique possibilities that psychical[1] phenomena open up for individuals to become actors in their lives but also the challenges that these phenomena's peculiarities impose on individuals in everyday life, especially for their possibilities to exchange with others. The various solutions that many species and, in particular, humans have evolved to overcome these challenges are explored along the microgenetic, ontogenetic and (possible) phylogenetic pathways in the development of individuals and the workings of their minds.

Know Thyself: The Intricate Challenges of Scientists Exploring Individuals

The TPS-Paradigm explicitly considers the fact that all science is made by human individuals and elaborates the particular implications that this fact entails. Specifically, all scientific endeavours depend on and reflect the particular perceptual and conceptual abilities of the human species—and their particular limitations. These human abilities determine the sole access that scientists can gain to the reality of the universe in which humans have evolved as a species. Therefore, TPS-Paradigm conceives of anything that is perceptible by the human senses (or that can be made perceptible, e.g. through technical means) and/or that can be conceived of by the human mind as a *phenomenon*. This notion differs from various historical research traditions in which phenomena are conceived of as mere sensory perceptions that are differentiated from non-sensual concepts (sometimes called noumena); for example,

[1] For the term psychical in differentiation to psychological, see part 13.4 below

in the philosophies of Plato or Kant (Hirschberger 1980a, b). The TPS-Paradigm explicitly refrains from making such distinctions for reasons explored below.

Human individuals are known to be prone to many kinds of biases, illusions, errors and fallacies in perception and reasoning (e.g. Wolpert 1992). Of particular importance for scientists of all disciplines are the fallacies that are derived from the human tendencies to seek regularities and structures and to oversimplify complexity, which are referred to as the *law of least effort* (Royce 1891). Of particular importance for scientists exploring humans is the *fallacy of misplaced concreteness* (Whitehead 1929) that is derived from the common tendency to uncritically assume that words correspond to concrete entities. But this may be possible only for words denoting physical events that can be directly perceived. It is not possible for words denoting abstract ideas that can only be conceived by human minds and that Bentham (1748–1832) referred to as "fictions" (Ogden 1932). But humans tend to assume that linguistic abstractions, "fictitious" words according to Bentham, can reflect real concrete entities.

Scientists exploring human individuals encounter further intricacies because they themselves are always individuals—and thus not independent from their objects of research. The scientists' own positions in their social world (unintentionally) influence the ways in which they explore individuals. In addition to the risks for introducing *anthropocentric biases* to their research that all scientists face, scientists exploring individuals are prone to introducing all kinds of *ethnocentric biases*, such as biases that are derived from their own language (Deutscher 2010), sociocultural and national background (Adam and Hanna 2012; Faucheux 1976; Russell 1927; Teo and Febbraro 2003), religion and worldview (Weber 1930, 1946), education and scientific tradition (Geertz 1988; Kuhn 1976), historical time (Fischer 1970; Gergen 1973) and age and gender (Pellegrini 2011). Finally, the scientists' own personal standpoints as individuals derived from their own personal experiences that they have made in their own lives entail additional risks for introducing all kinds of *egocentric biases* to their research (Fahrenberg 2013; Ramón y Cajal 1897/1999; Weber 1949).

Fallacies and biases influence not only the specific theories and models that scientists develop. More profoundly, they influence the very means by which scientists generate knowledge (Uher 2015a, b, c).

The Philosophy-of-Science Perspective: Scrutinising the Own Basic Assumptions

Knowledge about the making of science is the most general level of scientific knowledge and is referred to as *philosophy-of-science*. In philosophy-of-science, scientists make explicit and critically reflect on the philosophical presuppositions that they have made about the properties of the phenomena to be explored and about the fundamental notions by which knowledge about them can be gained (Aristotle 350 BCE; Fahrenberg 2013; Collingwood 1940; Toomela 2012; Walach 2013).

Philosophical presuppositions are basic rational structures that scientists conceive for a given scientific system and that are required for that system to func-

tion. Importantly, these rational structures can originate only from outside the given scientific system that is built on these structures; therefore, they cannot be rationally justified or validated within the system for which they are formulated (cf. incompleteness theorem; Gödel 1931) and are also called *absolute presuppositions* (Collingwood 1940). For example, many scientists presuppose that, in nature, there are basic structures that follow rational or logical laws and that can thus be described in logically consistent ways and be explored by means of rational or logical analyses. This presupposition appears to be self-evident, but actually it is not. Rather, scientists can also conceive alternative absolute presuppositions from which competing and contradicting scientific systems can be constructed (Fahrenberg 2013; Kellert 1993; Walach 2013).

On the basis of the particular absolute presuppositions being made, scientists develop metatheoretical and methodological frameworks that are coherent within the given scientific system yet not necessarily with the metatheories and methodologies used in other scientific systems that build on alterative absolute presuppositions. For example, psychologists have developed very different absolute presuppositions about human nature (the "images of man"; Fahrenberg 2004; Shotter 1975), such as the ideas that humans are driven by subconscious inner urges and conflicts (Freud 1915), are passively responding to external conditions (Skinner 1971; Watson 1913) or actively striving for cognisance (Kelly 1955) and personal growth and fulfilment (Maslow 1943; Rogers 1959). These different presuppositions laid the foundations for the development of various psychological research paradigms that each comprise coherent set of statements, theories and methods but that still tend to be contradictory or even irreconcilable with one another because they are based on different absolute presuppositions (Fahrenberg 2013; Walach 2013).

Making explicit the absolute presuppositions on which a given scientific system is built, therefore, is an essential prerequisite for analysing from a meta-perspective the theories, approaches and methods that are applied within a given system, thus for critically reflecting on the metatheories and methodologies that are used in a given field. *Metatheories* refer to the implicit and explicit beliefs, theoretical ideas and basic assumptions that scientists make about their objects of research and to the questions that they ask about these objects. The scientists' metatheories determine the ways in which they reduce real phenomena to scientific phenomena and thus, what they consider to be facts in their field and how the thus-defined facts can be theoretically analysed and interpreted (Althusser and Balibar 1970; Køppe 2012; Toomela 2011; Wagoner 2009; Weber 1949). *Methodologies* refer to the ways (i.e. approaches) in which scientists tackle the questions that they ask about their objects of research and to the techniques and research practices (i.e. methods) that they therefore use. Methodologies are closely interrelated and intertwined with the metatheories that scientists have derived from the particular philosophical presuppositions on which a given scientific system is built (Sprung and Sprung 1984; Uher 2013).

The TPS-Paradigm is called a *philosophy-of-science paradigm* not because it contains a philosophy-of-science; this is true for any scientific system. Instead, its name derives from its aim to make explicit as comprehensively as possible the absolute presuppositions, metatheories and methodologies on which it rests to en-

able scientists to critically reflect, discuss and further develop established theories, models and research practices and to derive ideas for novel lines of research. This is seldom done in the sciences exploring individuals (Fahrenberg 2013; Omi 2012; Schwarz 2014; Toomela 2011; Uher 2013, 2015a, b; Walach 2013; Weber 1949; Westen 1996).

Transdisciplinarity: Integrating and Applying Knowledge Across Disciplines

The TPS-Paradigm is *transdisciplinary* because, in contrast to other research paradigms, it explores concepts, approaches and methods that were developed in *different* established paradigms and different research disciplines studying individuals. By systematically elaborating the philosophical presuppositions, metatheories and methodologies on which different concepts, approaches and methods are built, the TPS-Paradigm identified commonalities and differences between them. This allowed for the coherent integration of concepts, approaches and methods from different disciplines into interrelated philosophical, metatheoretical and methodological frameworks that can be applied across disciplines. These frameworks also enabled the further and new development of concepts, approaches and methods that meaningfully complement and expand the existing ones. The TPS-Paradigm is targeted at supporting scientists to critically reflect, discuss and further develop previously established theories, models and research practices and to derive ideas for novel lines of research in the future (Uher 2011a, 2013, 2015a).

A comprehensive application of the TPS-Paradigm was demonstrated in "personality" psychology. By elaborating the particular metatheories and methodologies that scientists use to establish comprehensive "personality" taxonomies, the TPS-Paradigm revealed profound mismatches between the scientists' implicit and explicit metatheories and the methodologies applied. It was shown that comprehensive taxonomic models of individual-specificity in central phenomena explored in individuals, such as behaviours and experiencings, have not yet been developed. The application of the TPS-Paradigm also enabled the development of novel theoretical and methodological approaches that can fill the gaps identified and that can meaningfully complement and expand previous lines of research (Uher 2015b, c).

The Transdisciplinary Philosophy-of-Science Perspective Taken on Individuals as Living Organisms

Central to the TPS-Paradigm are the absolute presuppositions that are being made about individuals as living organisms and the metatheories and methodologies that are therefrom derived. In line with its transdisciplinary scope, the TPS-Paradigm builds on a broad array of theoretical concepts and methodologies from diverse

research disciplines from across the life sciences and beyond. Of particular importance are theoretical ideas about individuals as *living organisms*.

Living organisms can be conceived of as *systems* that are composed of sets of interrelated entities forming a complex whole (von Bertalanffy 1973). Living systems are complex at every level of their hierarchical structure. Entities at one level are compounded into *new* entities at the next higher level such that series of systems reside within a greater array of even more complex systems (e.g. nuclei, cells, organs, individuals, groups, communities, societies, species; Caprara 1996).

> Systems at each hierarchical level have two properties. They act as wholes (as though they were a homogeneous entity), and their characteristics cannot be deduced (even in theory) from the most complete knowledge of the components, taken separately or in other combinations. In other words, when such a system is assembled from its components, new characteristics of the whole emerge that could not have been predicted from a knowledge of the constituents. Such emergence of new properties occurs also throughout the inanimate world, but only organisms show such dramatic emergence of new characteristics at every hierarchical level of the system (Mayr 1988, p. 15).

The *whole* is not just more than the sum of its parts; it is essentially *different* from the sum of its parts—it has different properties, structures and functionings; this is referred to as the *principle of emergence* (Koffka 1935; Köhler 1969; cf. also Durkheim 1919; Simmel 1908). Hence, the identification of lower-level constituting elements of living organisms in and of itself cannot provide explanations of how the identified elements function together as a whole (e.g. Diriwächter and Valsiner 2008; Hartmann 1964; Pauli 1927; Koffka 1935; Köhler 1969; Toomela 2012; Vygotsky and Luria 1930; Wundt 1863). Moreover, as in living systems, series of systems are nested within each other, the entities that can be conceived at any one level can be conceived of as multi-contextual. Different properties and functionalities can emerge from the same set of elements in different contexts (Uher et al. 2013a; Uher 2015b; Walach 2013). Thus, the principle of emergence also entails that assumptions of isomorphisms between elements on different levels (in all directions) can be very misleading (Mayr 1988; Wolpert 1992). Isomorphisms are particularly low, if not completely absent, if phenomena of different kind—and thus with different metatheoretical properties—are concerned (see below).

As living organisms, individuals can be conceived of as *self-preserving* and *self-organising* from within their boundary (Luisi 2003; Varela et al. 1974; Zeleny 1977). Living organisms also exchange with their external surrounding and can therefore be conceived of as open systems (i.e. *dissipative systems*; Prigogine 1996). Dissipative systems develop non-linear system dynamics. In their developmental pathways, bifurcations may occur at which point the directions of the future development of a given system become unpredictable. In the development of living systems, dialectical processes occur in which interactions between elements can result in changes of the elements in and of themselves. These peculiarities of living organisms result in processes in their microgenetic, ontogenetic and phylogenetic development that are irreversible and historically unique (Baldwin 1896a; Caprara 1996; Li 2003; Prigogine 1996; Valsiner 2014).

Three Metatheoretical Properties that Determine the Perceptibility of Phenomena by Individuals

Central to the TPS-Paradigm are the absolute presuppositions that it makes about the phenomena explored in individuals, in particular about the differentiation of different kinds of phenomena from one another and the conception of their particular properties. The paradigm considers three metatheoretical properties that can be conceived for the various phenomena being studied. These metatheoretical properties are considered because the particular constellation of the forms that can be conceived for each given kind of phenomenon with regard to these properties determines the phenomena's perceptibility by individuals. Perceptibility by individuals has elementary consequences for the ways in which information from a given kind of phenomenon can be converted into information encoded in other kinds of phenomena (Uher 2015a, c). The central roles that such conversions play in individuals' everyday lives, such as for transmitting meanings, are explored in this research.

The absolute presuppositions that the TPS-Paradigm makes about the three metatheoretical properties and the distinctions between various kinds of phenomena need not be consensually shared by all scientists exploring individuals. Other scientist may make other absolute presuppositions and use other rationales to conceive of and to differentiate between phenomena and their properties. Those scientists who do not agree with the particular presuppositions made in the TPS-Paradigm must develop metatheoretical considerations other than the ones that are explored in this research, thus precluding direct comparisons. The explorations presented in this research are aimed at revealing possible differences in the absolute presuppositions that are made in the field and, in particular, to enable comparisons and controversial discussions between different research traditions and scientific disciplines that are based on the *same* absolute presuppositions as made in the TPS-Paradigm.

Importantly, the TPS-Paradigm generally considers the dimensions of everyday life experiences (i.e. spatial dimensions comparable to the human body, temporal dimensions of the international time standard) rather than to the dimensions of atoms or the outer space as considered in specific fields of research (e.g. chemistry, quantum physics, astronomy). The three metatheoretical properties, however, are conceived on levels of abstraction that are commonly not considered in either everyday life or science.

1. **The Phenomena's Location in Relation to the Individual's Body**
   The TPS-Paradigm considers the phenomena's spatial location in relation to the individual's body in terms of their *externality/internality*. Phenomena can be located internal or external[2] to the individual's body; some kinds of phenomena can also be both (e.g. body heat). The spatial location of phenomena has important consequences for individuals as it determines their opportunities to directly perceive these phenomena in themselves and in other individuals.

[2] The differentiation is made between internal versus external, rather than between *endogenous* versus *exogenous* because the latter implies a reference to potential causes and thus to explanations, which are not needed to metatheoretically define the various kinds of phenomena.

The TPS-Paradigm defines as *extroquestion* (from the Latin *extro* meaning beyond, outside and *quaestio* meaning seeking, investigation, enquiry) the exploration of phenomena that individuals can directly perceive as being located external to their bodies, which allows multiple individuals to perceive one and the same event. Joint perception is important because it facilitates intersubjective perception and social exchange as explored below. Importantly, extroquestion is defined on the basis of the phenomena under study and of the individuals who are perceiving these phenomena and providing information about these phenomena from their pertinent perceptions and conceptions (Uher 2015a). Extroquestively accessible are all phenomena that are external to individuals' bodies, thus physical ones (e.g. individuals' physiognomy, written language, monuments). Through the use of invasive methods (e.g. endoscopy, surgery), physical phenomena that are internal to individuals' bodies (e.g. morphology of inner organs) can also become perceptible by multiple individuals.

Other internal phenomena, by contrast, cannot be perceived by other individuals at all; they can be accessed only by the single individual. These are the unique properties of the phenomena of the psyche (e.g. experiencing). The TPS-Paradigm defines as *introquestion* (from the Latin *intro* meaning in, within) the exploration of phenomena that can be directly perceived only by the individual him- or herself and that are, in principle, not directly perceptible by any other individual under all possible conditions—thus, psychical phenomena (see below; Uher 2015a).

Extro*spection* and intro*spection*, by contrast, are commonly defined and differentiated from one another on the basis of the *perspective* that individuals can take on themselves versus on other individuals or on things. However, both perspectives are always interwoven as individuals can always extrospect and introspect at the same time (cf. Kant 1781; Wundt 1896).

2. The Phenomena's Temporal Extension

The TPS-Paradigm considers the phenomena's temporal extension because individuals can directly perceive only those phenomena that are *present in a given moment*. Some phenomena are temporally more extended and persist over some period of time (e.g. individuals' body morphology). This facilitates the phenomena's direct perception by individuals in their everyday lives. Other phenomena are much less temporally extended and change more quickly (e.g. blood sugar, hairstyle). Still other phenomena, in turn, are strictly momentary and highly fluctuating (e.g. behaviours, thoughts); their occurrence is strictly bound to the present moment in time—the here and now. Momentary phenomena can be directly perceived only in the very moments in which they occur (e.g. kicking a ball) or when they have caused changes in other phenomena that can still be directly perceived (e.g. the ball lying in the goal area).

3. The Phenomena's Physicality versus "Non-Physicality"

The TPS-Paradigm considers the phenomena's *physicality* because material physical phenomena feature *spatial units* that are rather constant and identically

repeatable to a considerable extent (e.g. electrons, atoms, molecules). Such spatial units help in creating an intersubjective consensus between individuals on how to categorise and compare the phenomena and events encountered in life. Spatial units occur in the material phenomena of individuals' bodies (e.g. cells, organs, body parts) and in the material physical phenomena of their external surroundings (e.g. other individuals, plants, animals, printed texts, buildings). Material physical phenomena can also be used to determine rather constant and repeatable units in immaterial physical phenomena, which feature no spatial units in and of themselves, such as behavioural and many physiological phenomena (e.g. body movements, body heat). Importantly, in the TPS-Paradigm, the terms physical and physicality refer to the science of physics and not to corporality, which cannot be conceived for immaterial physical phenomena.

The phenomena of the psyche, in and of themselves, feature properties that are of an entirely different kind than the properties of physical phenomena and that are therefore conceived of as "non-physical" in the TPS-Paradigm. The term is put in quotation marks because it does not indicate a simple contrast to physical. Rather, the term is meant to denote properties that are often associated with the terms psychical and mental and also with intangible and immaterial. But as behavioural and some physiological phenomena can also be conceived of as being immaterial, the paradigm uses the term "non-physical". The term denotes that psychical phenomena, in and of themselves, are immaterial and that spatial units that are identically repeatable, at least to some degree, cannot be found. But in contrast to immaterial physical phenomena, there are no systematic relations between psychical phenomena and the (material and immaterial) physical phenomena that accompany them (e.g. chemical and electric phenomena in the brain; Fahrenberg 1979, 1992, 2008a, b; Kant 1798; Wundt 1894).

The Epistemological Principle of Complementarity and the Psyche-Physicality Problem

The differentiation of physical and "non-physical" properties as made in the TPS-Paradigm refers to one of the most fundamental research problems in philosophy and psychology—the so-called body–mind problem or brain–mind problem (e.g. Fahrenberg 1979, 1992), which is referred to as the *psyche-physicality problem* in the TPS-Paradigm, in line with its particular terminology. The absolute presuppositions that the TPS-Paradigm makes about this problem are based directly on the *principle of complementarity* introduced by Bohr (1937) in quantum physics as a solution for the wave-particle dilemma in research on the nature of light. This epistemological principle considers the fact that, in living and non-living nature, pairs of properties can often be found that are mutually exclusive and maximally incompatible with one another but that are both related to the same object of research and both necessary for its sufficient description. Consequently, the different metatheoretical properties that can be conceived for the phenomena explored in individuals

and the peculiarities that they entail for the phenomena's direct perceptibility by individuals (e.g. researchers) must be adequately considered both in metatheoretical explorations (see below) and in the research methodologies used for investigations (see Uher 2016; for applications in philosophy and psychology, see e.g. Fahrenberg 1979, 1992, 2013; Hoche 2008; Kant 1798; Uher 2015c; Walach 2013; Wundt 1894).

The TPS-Paradigm conceives only of physical and "non-physical" properties as being complementary to each other in the Bohrian sense. But the different forms that phenomena can take in each of the two other metatheoretical properties (i.e. internal/external and momentary/not momentary) are not conceived of as complementary because each of these two properties can be conceived of as reflecting a gradual dimension representing the same kind of property (i.e. location in relation to the individual's body, temporal extension). The three metatheoretical properties in and of themselves are not complementary to each other either. By contrast, each given kind of phenomenon is *always* characterised by a particular *constellation of forms of all three properties*, allowing for the metatheoretical differentiation of different kinds of phenomena.

The Various Kinds of Phenomena Differentiated in the TPS-Paradigm

On the basis of the three metatheoretical properties and the particular constellation of their forms that can be conceived for a given phenomenon, the TPS-Paradigm differentiates various kinds of phenomena explored in individuals.

Two major groups are conceived, basic kinds and composite kinds of phenomena. The phenomena of morphology, physiology, behaviour and the psyche are conceived of as *basic kinds of phenomena* because they cannot be removed from the body of the individual being considered without destroying its integrity (Uher 2015a). By contrast, the phenomena of semiotic representations, artificially modified outer appearance and contexts ("environment") are conceived of as *composite kinds of phenomena* because they each comprise several different kinds of phenomena, among them at least one basic kind of phenomenon, which is thus inseparable (in the sense stated) from the body of the individual considered. Composite kinds of phenomena may also comprise external physical phenomena that may be bound to or independent from the individual being considered. Hence, composite kinds of phenomena comprise phenomena with heterogeneous metatheoretical properties, which entail that their structures and interrelations are highly complex. The basic kinds of phenomena, by contrast, as they comprise only one kind of phenomenon as differentiated in the TPS-Paradigm, have comparably homogeneous metatheoretical properties, and their structures are therefore less complex (see below).

The following sections explore these various kinds of phenomena, their particular constellations of metatheoretical properties and the implications that these constellations entail for the phenomena's perceptibility by individuals in their everyday lives. Some phenomena are discussed only briefly (e.g. morphology, physiology).

The main focus is on explorations of psychical phenomena and of those kinds of phenomena that comprise psychical phenomena or that have important functional interrelations with psychical phenomena in individuals' everyday lives.

13.2 The Individuals' Body: Morphology and Physiology

Morphology refers to the structures and constituting parts of individuals' bodies in the TPS-Paradigm. Morphological phenomena can be located both internal to the individuals considered (e.g. morphological brain structure, skeletal build) and external to them (e.g. physiognomy); some phenomena can be both (e.g. hair). Morphological phenomena are temporally extended; they can and do change over time, in particular during ontogeny, but they change only slowly and are thus more persistent. Morphological phenomena are material physical; one and the same event can therefore be directly perceived by multiple individuals. Many external phenomena of individuals' morphology (e.g. physique) can be directly perceived with the naked eye or bare hands by individuals in ordinary everyday life settings. This is not possible for external micro-level phenomena (e.g. cell surfaces of the outer skin) and all internal phenomena of morphology (e.g. intestinal structures, except in accidents). But, in present-day humans, they can be made perceptible by using invasive methods (e.g. surgery) and technical means (e.g. microscopes, endoscopes). In addition, morphological phenomena, because they are material physical, feature spatial units that are identically repeatable to considerable degree (e.g. molecules, cells, organs, body parts). Together with their extroquestive accessibility, this facilitates reaching intersubjective consensus between individuals on how to categorise phenomena and events. This also allows for direct comparisons within and between individuals and with designated spatial standards of measurement (e.g. metering rule), thus enabling scientific quantifications (Uher 2013, 2015a, 2016).

Physiology refers to the functioning of the morphological structures of individuals' bodies in the TPS-Paradigm. Physiological phenomena are primarily located internal to the individual (e.g. neurotransmitter systems), but some can also become external (e.g. heat). Most physiological phenomena are not bound to the immediate moment, but their temporal extension varies from phenomena that occur only briefly (e.g. motor unit action potentials) to phenomena that are more persistent (e.g. blood circulation, body temperature). Some physiological phenomena can be conceived of as material (e.g. chemical signals). Others are immaterial though bound to the individual's bodily matter, which facilitates the identification of spatial units on which categorisations can be based (e.g. breaths, heartbeats). The physical properties of physiological phenomena also enable direct comparisons within and between individuals and with designated physical standards (e.g. barometer to measure blood pressure), thus enabling scientific quantifications (Uher 2013, 2015a, 2016). Given this constellation of metatheoretical properties, individuals can directly perceive in other individuals only a few physiological phenomena in everyday live situations (e.g. sweat). For most physiological phenomena, individu-

als must employ invasive methods (e.g. blood sampling) and use technical means (e.g. stethoscope). But within themselves, individuals are able to sense and perceive some of their own physiological phenomena as explored below.

13.3 Behaviours: Individuals' Primary Way to Connect with External Phenomena

The TPS-Paradigm metatheoretically defines behaviours as "external changes or activities of living organisms that are functionally mediated[3] by other external physical phenomena (Millikan 1993) in the present moment" (Uher 2013, 2015a; Uher et al. 2013a, b). External changes and activities can be mere by-products of the organisms' chemistry (e.g. heat) and physics (e.g. breath sounds) or they may fulfil functions of the organisms' physiological regulations (e.g. loss of heat serving thermoregulation). Therefore, not any functional externalisation or external change can be conceived of as behaviour; they are behaviours only if their *functions have reference to other external physical phenomena* or relations to them (Millikan 1993). This metatheoretical definition implies that, to identify the function of a particular behavioural event, the external contexts in which it occurs must be considered (e.g. raising an arm to reach a fruit in the tree or to threaten an opponent; see the behavioural situation, below).

This metatheoretical definition of behaviour differs in important ways from previous concepts in psychology. Specifically, the TPS-Paradigm conceives neither of physiological responses nor of mental activities as behaviours; this differs fundamentally from behaviouristic concepts (e.g. Skinner 1957). Instead, it explicitly considers that physiological and mental phenomena have different metatheoretical properties and it therefore conceives of them as constituting different kinds of phenomena. For the same reason, the TPS-Paradigm refrains from making a priori assumptions about the potential causation of behavioural phenomena in other kinds of phenomena as is implied, for example, by the concepts of behavioural "responses" or "goal-directed" actions. On the basis of the absolute presuppositions made about individuals as living organisms, it explicitly considers that events of a given kind of phenomenon can be dynamically interrelated to and co-determined by events of *all* other kinds of phenomena in various (subsidiary) systems both internal and external to the particular individual considered. In line with another important absolute presupposition made in the TPS-Paradigm—the Bohrian principle of complementarity—each given kind of phenomenon is first defined in its own right concerning its own particular constellation of forms with regard to the three metatheoretical properties considered. This is an essential prerequisite for the selection and development of methodologies that enable appropriate investigations and for the analysis of pos-

[3] In the TPS-Paradigm, the term *mediation* refers to the Latin *mediare*, to be in the middle, not to the meaning of mediation as used in statistics (where it is differentiated from moderation).

sible causal interrelations between events of different kinds of phenomena (Uher 2015a, b, c).

Hence, in the TPS-Paradigm, all behavioural phenomena are conceived of as being located external to individuals' bodies. Moreover, behavioural phenomena are bound to the immediate moment; their events are ephemeral and highly fluctuating. Behavioural phenomena are also bound to or emanate from the individuals' bodies; but, in and of themselves, they are immaterial physical phenomena (e.g. movements, acoustic waves). Behaviours are continuous and dynamic processes in which spatial units suggesting clear demarcations of single events are largely absent. But demarcations can be made based on the material physical properties of the individual's body to which they are bound, which also entails the identical repeatability of events to some extent (e.g. events of scratching can be demarcated through finger flexions).

The external and physical properties of behaviours enable multiple individuals to directly perceive one and the same event, which facilitates finding intersubjective consensus on how to demarcate and categorise events—but within the constraints of the behaviours' lack of spatial units. For example, what is one event of a scratch ($n = 1$) given that finger flexions can differ in both extension and speed and that individuals can use one or multiple fingers or even both hands? The constraints of the behaviours' limited temporal extension further complicate individuals' possibilities to jointly perceive one and the same behavioural event because individuals can perceive behavioural events only *while* they happen (e.g. a hug, a smile) or while they are still ongoing (e.g. bouncing, running). The momentariness of behavioural events also complicates direct comparisons with designated spatial standards for enabling scientific quantifications of events (e.g. the loudness of a sound can be perceived only while it occurs). As behavioural phenomena are external and physical, these constraints can be reduced to some extent by technically converting information from these phenomena into information in other kinds of physical phenomena (e.g. audio records; see Uher 2015c)—at least, this is possible for present-day humans. The momentariness of behaviours also complicates comparisons of events displayed by the same individual (except for concurrent events) and comparisons of events displayed by different individuals. Because behaviours are momentary and occur seldom spatially and temporally exactly in parallel with one another, individuals can compare ongoing behavioural events only with past events, which necessarily have already ceased to be and of which individuals can retain only memories. But memories are different kinds of phenomena than behaviours—they are phenomena of the psyche.

13.4 The Psyche: The Individual's Inner World

In the TPS-Paradigm, the psyche denotes the entirety of the immediate experiential reality both conscious and non-conscious of living organisms—the individual's inner activity and inner world (*Innenwelt*, von Uexküll 1909).

The TPS-Paradigm builds on concepts of the psyche that are rooted primarily in German-language philosophy and psychology from the nineteenth and twentieth century—from the research areas and times in which the science of the psyche—psychology—was established as a scientific discipline. To use, integrate and elaborate these concepts, the paradigm introduces some terminological differentiations that are commonly not made in the pertinent English-language literature. Contrary to common practice, the paradigm translates the German term *psychisch* into *psychical*[4] rather than *psychological* (German: *psychologisch*) because "events, processes, and structures that are properly called psychical do not become psycho*logical* until they have been operated upon in some way by the science of psychology" (Adams and Zener in Lewin 1935, p. vii; emphases added). A further differentiation made between experiencing and experiences is explained below.

Importantly, the TPS-Paradigm considers all kinds of psychical phenomena (e.g. those commonly referred to as thinking, feeling, wanting, etc.) rather than focussing on only particular ones (e.g. only thinking) because all these phenomena share the same constellation of the three metatheoretical properties considered in the TPS-Paradigm and thus, cannot be differentiated on the basis of these properties. Moreover, individuals' immediate experiential reality always comprises all kinds of psychical phenomena (Wundt 1896). For these reasons, the phenomena of the psyche are primarily referred to as psychical rather than as mental because the term mental is often used to refer to cognitive phenomena only, thus excluding emotional, volitional and other kinds of psychical phenomena. Considering all kinds of psychical phenomena is important for holistic explorations of individuals that are in the focus of this research.

Psychical phenomena belong to the phenomena of life. As such, they are bound to a unit—the individual (Pauli 1927). Thus, they are also bound to and directed toward the individual's life (Stern 1924, p. 203). Er*leben* (experiencing) presupposes Leben—life. This is reflected in the term *psyche* originating from the ancient Greek word *ψυχή* for life, breath. As psychical phenomena are bound to the individual organism, each psychical event is dynamically interrelated to and co-determined by all concurrent events and by past events within the same organism (Lewin 1935). As a consequence, no single event can be conceived independently from all other events, and interrelationships between psychical phenomena are highly complex (Pauli 1927; Rothschuh 1963). In addition, like all phenomena of life, psychical phenomena vary intra-individually and inter-individually. Psychological laws therefore cannot be deterministic but only probabilistic (Brunswik 1952, 1955; Uher 2013)—unlike many natural science laws describing phenomena of non-living matter (Pauli 1927; Schrödinger 1944).

As phenomena of life, psychical phenomena are directly and intimately interrelated with the individual's organismal processes of life (Lewin 1935; Pauli 1927;

[4] Similarly, people are allergic, not allergologic; it is the medical treatments of allergies that are allergologic and that are developed by the science studying allergies, allergology. Unfortunately, such differentiations are not made consistently in either English or German; for example, biological (biologisch) refers to both the organisms' phenomena and their scientific exploration.

Schrödinger 1958). The emergence of psychical phenomena essentially presupposes and is bound to physical phenomena of life. Psychical phenomena, in and of themselves, are immaterial; but in contrast to all other phenomena of life, they cannot be conceived of as being physical. The properties of psychical phenomena essentially differ from material and immaterial physical phenomena because they feature no spatial units or at least rather constant interrelations to the physical phenomena (e.g. electric and neurotransmitter activity in the brain) that accompany them (Fahrenberg 1979, 1992, 2008a, 2008b; Kant 1798; Wundt 1894). These properties are conceived of as "non-physical" in the TPS-Paradigm. They entail that the "psyche", as the entirety of psychical phenomena, cannot be conceived of as a spatial entity that could be directly perceived (as is possible for the individual's body); it therefore does not and cannot imply reification as a concrete entity. The psyche can only be conceived of as an entity, thus as a subjectively or intersubjectively constructed entity (Uher 2015a).

In contrast to all other phenomena of life, psychical phenomena are entirely internal and directly accessible only by their carrier (Pauli 1927), thus introquestively, and they are inherently subjective and idiosyncratic (Weber 1949). One and the same event can never be perceived by multiple individuals (Locke 1689), precluding direct comparisons of their phenomenal properties (called *qualia*; Levine 2003) between individuals (Schrödinger 1958; Toomela 2008; Uher 2013). Internality and "non-physicality" of psychical phenomena entail particular intricacies—both for their carriers and for other individuals.

As phenomena of life, psychical phenomena are interrelated also with events that are external to the individual (Brunswik 1952; Lewin 1935). External physical events can directly interact with events in the individual's psyche through sensation and perception (see below) and the internal physical phenomena with which psychical phenomena are connected in complementary ways (e.g. sensory organs). But, vice versa, psychical events in and of themselves and the internal physical phenomena accompanying them cannot directly connect with external phenomena and thus cannot have any direct effect on external events (Schrödinger 1958; Sherrington 1940). This is referred to as the *one-sided psyche-external surrounding connection*[5] in the TPS-Paradigm. Bridging this one-sided gap requires externalisations—other kinds of phenomena that serve as mediators from the individual's internal physical and psychical phenomena to external physical phenomena (Uher 2013, 2015a).

The individuals' primary mediators for externalising information from psychical events are behaviours, including behavioural events that form part of semiotic language (see below; Uher 2013). The morphological and physiological phenomena that are functionally necessary for behavioural phenomena to occur (e.g. muscle fibres and their enervation) are not specifically considered because they are internal to the individual's body and thus, cannot in and of themselves directly connect to phenomena in the individual's external surroundings. Behavioural phenomena are so flexible and so neatly intertwined with psychical phenomena that individuals hardly notice the behaviours' mediating function in externalising information from

[5] Previously labeled the *mind–environment connection* (Uher 2013).

psychical phenomena. This may contribute to conceptions of psychical phenomena as "inner behaviours" (e.g. Koffka 1935; Skinner 1957; Sprung and Sprung 1984). The philosophy-of-science perspective taken in the TPS-Paradigm highlights that for behavioural phenomena different constellations of metatheoretical properties can be conceived than for psychical phenomena. These different constellations entail profound differences in these phenomena's perceptibility by individuals and therefore require conceptual differentiations (Uher 2015a).

Importantly, the different metatheoretical properties that can be conceived for the phenomena that individuals use for externalisations precludes one-to-one conversions of information from psychical events. Isomorphisms between phenomena of different kind are generally low, if not largely absent (Wolpert 1992). This has important implications for individuals' possibilities to make inferences from behavioural events to psychical events. Moreover, externalising phenomena are connected not only with psychical phenomena but also with further kinds of phenomena, both internal and external to individuals. These interrelations influence these externalising phenomena in ways that are unrelated to the psychical phenomena from which information is being externalised, which further constrains individuals' possibilities for unequivocal externalisations and, vice versa, for making inferences from behavioural events to psychical ones (Uher 2013, 2015a).

In a nutshell, psychical phenomena can be conceived of as multi-contextually embedded into the individual's life, both internally and externally. The TPS-Paradigm therefore refers to the psyche also as the individual's psychical system, implying some properties of organisation that are common to all systems of living beings. Like all living systems, psychical systems can be conceived of as self-organising and therefore also as self-referential to considerable extent (cf. Luisi 2003). As living systems, psychical systems can be explored for both their compositional structures and the process structures by which their structural components function together in continuous and irreversible ways of development (Caprara 1996; Sato et al. 2010; Uher 2015c; Valsiner 2000, 2012). The speed of change and development essentially differs between two kinds of psychical phenomena that the TPS-Paradigm differentiates on the basis of their temporal extension.

Experiencings and Memorised Psychical Resultants

In line with concepts and terminology from German-language psychology, the TPS-Paradigm differentiates two kinds of psychical phenomena. Contrary to common practice in the English-language literature, the German term *Erleben* (Stern 1924) is translated as *experiencing* that is opposed to the *experiences,* German *Erfahrungen,* that one can make in terms of information gained from past events of experiencing. Erleben and Erfahrung both translate into experience; but they are not the same. Erfahrung is derived from Erleben; it is the empirical—the a posteriori—whereas experiencing is bound to the immediate moment (see below). Therefore, empirical sciences are also called *Erfahrungs*wissenschaften in German; only few of these sciences explore experiencing in and of itself.

Experiencings are strictly bound to the present moment (Valsiner 1987, 2012). Experiencings are "actualities" (Gillespie and Zittaun 2010, p. 72), which Stern (1924) referred to as the "immediacy of the product of internaling[6]" (Unmittelbarkeit des Innerungserzeugnisses, p. 203). Pauli (1927) similarly ascertained experiencing is nothing persistent (nichts Beharrendes) but in continuous processes of change and characterised as ongoing events (als Geschehen). In this continuous flow of experiencing, every event is unique and never repeatable (James 1890; Salvatore et al. 2010; Toomela and Valsiner 2010; Valsiner 2012).

Events of experiencing leave "impressions" in the individual that change his or her overall psychical system (von Uexküll 1909).

> Animals' vital activities toward outer stimuli do not simply proceed as in any machine, the construction of which cannot change. In contrast, animals' body plan continuously changes under the influence of the surrounding, such that one can say with exaggeration, a stimulus never encounters the same animal twice[7] (von Uexküll 1909, p. 25).

Thus, individuals can retain in their psychical systems past events of experiencing in processed forms that are conceived of as *memorised psychical resultants of past experiencing* in the TPS-Paradigm. Importantly, events of experiencings are not simply stored in the same form as originally experienced. Experiencings are processed, thus becoming experiences that are interconnected with other experiences and integrated into the individual's psychical system the structure of which thereby continuously changes and thus develops (e.g. Le Poidevin 2011; Peirce 1902, CP 2.84; Sato et al. 2010; Uher 2013; Valsiner 2012). It follows that individuals can operate—internally and externally—only from within the repertoire of their hitherto reached systemic structure. This psychical structure, because it is *memorised*, is not strictly bound to the immediate moment—in contrast to the experiencings from which it results. Psychical resultants are inherently more temporally extended and therefore conceived of primarily as structures, although they are—just as experiencings, but necessarily slower than them—in continuous processes of development as well (Uher 2015a, 2015c; Valsiner 2000, 2012).

Not only is the genesis of memorised psychical resultants inherently bound to experiencing but also their use. Reviving an *Innerung* one had is an Er*innerung*—a remembering and re*minding* (Stern 1924). But reviving a past experiencing is not that same experiencing anymore because it has already ceased to be (Le Poidevin 2011; Uher 2013). Rather, it is a new experiencing that is (re)constructed in the given moment (Bartlett 1932) from the processed memory of that past experiencing as it has been retained and integrated in the hitherto reached mnemonic structure of

[6] "Innerung" is not listed in German dictionaries in contrast to its opposite Äußerung, which can be translated as expression or externalisation for which, however, separate German words exist (Ausdruck and Externalisierung). Therefore, "Innerung" is translated here in the likewise non-existing English word "internaling" rather than as impression or internalisation, for which separate German words exist as well (Eindruck and Internalisierung).

[7] In the original: "Es läuft die Lebenstätigkeit der Tiere auf äußere Reize nicht einfach ab, wie in irgendeiner Maschine, deren Bauplan sich gar nicht verändern kann. Im Gegenteil ändert sich der Bauplan der Tiere dauernd unter dem Einflusse der Umgebung, so daß man mit Übertreibung sagen kann, niemals trifft ein Reiz zum zweiten Male das gleiche Tier."

the individual's psychical system (cf. Schacter and Addis 2007). Retrieved and reconstructed experiencings are processed again (e.g. rebuild, remodelled, reshaped) before they are memorised anew—a fact well known in psychotherapeutic research (Kelly 1955). Experiencing is the working mode of psychical systems. Experiencing can be conscious and also subconscious (Freud 1915); but it ceases to be in various states of loss of consciousness, such as during deep sleep, anaesthesia and in some clinical conditions (e.g. vegetative state; Casali et al. 2013).

Because memorised psychical resultants can be retrieved only in experiencings, individuals commonly do not notice a sharp division between them as can be made on metatheoretical levels. This can be illustrated by the example of perceiving (Gibson 1967).

Sensory, Perceptual and Psychical Representations

Sense organs have evolved in ways that enable individuals to physically interact with physical events of particular kind. Interactions with events that are external to individuals are of particular importance for individuals' abilities to preserve their physical organismal properties. These physical interactions between external physical events (e.g. light) and individuals' sense organs (e.g. photoreceptor cells in the retina of the eye) produce neural signals—*sensations* (Gibson 1967; Schrödinger 1951). Sensory phenomena are internal to the individual's body; they are *sensory representations* of those external physical phenomena that elicit them. Like these external phenomena, sensory representations are also physical—some are material (e.g. chemical signals), others are immaterial (e.g. electrical signals). Sensory representations, given their different metatheoretical properties, internally present information from external physical phenomena in forms that differ from those of the phenomena from which information is being represented (e.g. the image of a tree that is created on the retina has different properties than the tree in and of itself that is being reflected and that is located external to the individual's body).

Sensations are physiological processes; but they are special ones operating at the border from the physical to the psychical into which they become processed as *perceptions* (Ader 2006). Sensory phenomena enable individuals to convert information from external physical events into information in internal psychical events—i.e. to externally perceive. Importantly, the patterns according to which sensations are converted into percepts are not fixed (see below) and sensations are not the only ways in which perceptions are generated (Gibson 1967).

Sensory impressions are occasional, highly fluctuating and incomplete because stimulus patterns in external events are never unchanging—already due to activities of the individual him- or herself, such as eye blinks or own movements. In addition, individuals' can flexibly shift their "perceptual lenses" and focus on particular details of physical events. This enables individuals to increase their sensory input but cannot fully make up for its inherent fragmentation (Brunswik 1952, 1956).

When individuals explore the invariant elements of an external physical phenomenon (e.g. by walking around a tree or by turning some of its leaves), variants in the individuals' sensations result from their own body movements. Because these changes have subjective reference to the individuals' own bodies (i.e. are *proprio-specific*; Sherrington 1906), these sensory variants can be controlled by the individuals themselves, thus enabling them to extract those sensory invariants that reflect properties of the external physical phenomenon that they are exploring. When invariant properties of external physical phenomena correspond to invariants in individuals' sensory impressions, individuals can obtain from these invariant sensory impressions information about the external phenomena under exploration and can develop *perceptual representation* of them (Gibson 1967).

Perceptual representations, unlike sensory ones, represent external physical phenomena in "non-physical" ways, thus *uncoupled from the physical laws* to which the physical phenomena that are being represented are bound. This further reduces the possibilities for isomorphisms between individuals' perceptual events and the physical events that are being perceptually represented.

Orientation toward life and self-maintenance implies that individuals can identify interrelations between elements—*information*—that are significant for their survival (cf. Gibson 1967). Individuals can psychically represent these interrelations as psychical associations. Of particular significance for living organisms are interrelations in and with the physical world. In millions of years, exposed to the physicality of this planet, many complex species have evolved that are equipped with organismal properties, enabling each of their individuals to develop psychical representations that are sufficiently functional for their survival in those particular details of the physical universe to which their species has adapted (cf. Darwin 1859; Merleau-Ponty 1967).

During ontogeny, individuals' sensory and psychical representations develop from tight interplays with their external surroundings, often promoted by active explorations that are characteristic for the young individuals of mammalian species. Changes in individuals' sensory representations are often confined to particular temporal windows during their ontogenetic physical development (Rosenzweig et al. 1999). But changes in individuals' psychical representations are, given their "non-physical" properties, theoretically unlimited.

By comparing with one another psychically represented elements and their associations, individuals can identify commonalities and differences. These psychical processes enable *abstractions, generalisations and categorisations*, thus internal organisations of the elements that are being psychically represented. Abstracted and generalised representations, because they are psychically derived, need not have direct counterparts in the external physical events that are being internally represented and analysed. No single tree exactly features the average properties that can be abstracted from many trees. Abstractions and generalisations are ideals—*ideas*—that represent only in approximated form those particular properties of external physical events that are important for the given individual (cf. Lahlou 1998).

In individuals' psychical systems, the *per*ceived—the what-is-taken-in—becomes *con*ceived—taken together in *concepts*. The pertinent everyday terms of the French language—*aper*cevoir (to perceive)–*con*cevoir (to conceive)–*voir* (to see)—direct-

ly reflect that individuals can "see" (i.e. visually perceive) only when the sketchy sensory input is put together. Concepts also enable individuals to perceive different physical events as being *of the same kind* (e.g. "trees"), although the single events occurring in individuals' sensory perceptions are never identically repeated (Brunswik 1956). Concepts enable individuals to perceive physical objects as stable despite the fact that individuals' sensations of physical objects are always incomplete and vary rapidly (Gibson 1967). These properties enable individuals to perceive events with just minimal sensory input—a glance becomes sufficient to "see" a "tree".

Individuals can also perceive internal physical events for which they are receptive, at least to some extent. As the events to be perceived occur internal to individuals, information from them need not be taken in from the external surroundings. Internal perceptions can also occur through bodily organs other than sensory ones (e.g. blood sugar levels). Individuals can process and abstract information from these internal perceptions, thus developing conceptual representations of inner organismal conditions (e.g. hunger).

Through these processes, individuals' psychical representations can become ever more detached from mere perceptual concepts. From *concrete* concepts, individuals can derive *abstract* concepts that no longer refer to concrete physical events that are directly perceivable. Individuals' conceptions of phenomena become independent from the embodied perception of single events of these phenomena. This enables individuals to internally represent the physical phenomena encountered in life also in the moments and situations in which the particular phenomena that are being psychically represented are not present (cf. Tomasello 2014).

Over time, psychically represented elements and associations and their abstractions and generalisations are taken together in concepts in ever more complex ways. Associations between concepts emerge resulting in networks of interconnected and contextualised concepts—i.e. *knowledge*. With increasing complexity and degrees of abstraction, new structures and qualities can emerge (cf. principle of emergence). By *per*forming (i.e. changing the forms of) psychical representations of physical phenomena, individuals can also infer properties and interrelations in the properties of physical phenomena that are not directly perceivable in and of themselves (cf. Tomasello 2014).

But inferences, as they are derived from psychical—thus "non-physical"—operations, are prone to the many fallacies, biases, illusions and errors that are intrinsic to human minds (Uher 2015a, c). Common-sense beliefs therefore represent the properties of physical matter that are important in individuals' everyday life in ways that are *viable* for the individuals' functioning in their particular physical surrounding (cf. Kelly 1955; Valsiner 2000). But regardless of their viability (e.g. in everyday life), individuals' psychical representations need not adequately *correspond* to what is given in physical phenomena and need not be *correct* in the scientific sense. In fact, physical laws are often ill-represented in everyday thinking (Wolpert 1992). Incongruencies between information in physical phenomena and information in individuals' pertinent psychical representations can become apparent when individuals notice (e.g. from exploration or experimentation) that their

psychical representations cannot predict the occurrences of physical events sufficiently well for enabling particular functionalities in interactions with these events. If such incongruencies are noticed and considered to be significant by (particular) individuals (e.g. scientists), given the abilities of psychical systems to self-organise, these psychical representations tend to approximate the new evidence, thus generating new knowledge.

Psychical Representations of Space and Time and of "Non-physical" Phenomena

The abilities to internally represent physical phenomena disembodied from their immediate perception and to process and abstract psychical representations enable individuals to identify between the physical phenomena that are being represented also spatial relations that are not directly perceivable in and of themselves. Individuals can psychically represent such abstracted spatial relations in cognitive maps that enable them to psychically identify novel ways that they have never used before to reach places that they already know (Haas 2004; Tolman 1948). With increasing psychical capacities, individuals can develop more general and more abstract concepts of *space*.

When individuals' psychical systems reach certain degrees of complexity and abstraction, individuals can also notice changes in the psychical representations that they have developed about the same physical phenomena. Individuals may notice, for example, that they psychically represent the sky as bright but also as dark or the selfsame tree with green leaves, but also with yellow leaves and without any leaves. Changes that individuals become aware of are conceived of as *time* (St. Augustine 397 CE). Like all experiencings, individuals' abilities to consciously perceive changes are embedded into the particular contexts of their current situatedness (see below). Therefore, individuals' abilities to become aware of changes vary within and between individuals—and thus also their perceptions of time (Le Poidevin 2004; Mellor 1985).

Awareness of physical changes that are directly perceivable in consecutive moments, such as the grains of sand in an hourglass flowing from the upper compartment into the lower one, leads to awareness of the passage of time in the *present* and of its flowing into the memories that result from these experiencings, which are conceived of as the *past*. By mentally projecting changes into the what-is-not-yet, individuals conceive the *future* (James 1890; Le Poidevin 2004, 2011; Valsiner 2012). Hourglasses illustrate this concretely: The upper compartment is filled with the grains of sand that have not yet passed (the future) the minuscule border through which some grains are continuously flowing (the present) into the lower compartment that thereby becomes steadily filled with those grains that have already passed (the past).

Thus, although time itself is real, tense is not (Mellor 1985); past and future are constructions of the human mind (St Augustine 397 CE). Time always flows in just

one direction (Prigogine 1996; Valsiner 2014). But individuals' psychical abilities, especially their mnemonic abilities, enable them to mentally travel backward and forward in time. Individuals use their memories of past events to imagine possible future events. This is also reflected in patterns of physical brain activity; the brain regions that are active when individuals retrieve past events and those that are active when individuals imagine possible future events show remarkable overlaps (Addis et al. 2007).

The ability to conceive of various tenses presupposes conscious awareness of the passage of time—an ability that develops only with increasing levels of complexity of individuals' psychical systems during ontogeny (Fraisse 1964; Piaget 1969). But conceptions of the passage of time are immanent to all psychical phenomena, even if not consciously noticed by individuals. In fact, the ability to conceive temporal relations is central for individuals' ability to extract invariants from their rapidly varying sensations, thus for developing perceptual representations and other kinds of memorised psychical resultants. Individuals may not notice this because they can retrieve and reconstruct memorised psychical resultants only in their experiencings in which the present merges indistinguishably with the past (Gibson 1967).

Individuals whose psychical systems have reached certain levels of structural and organisational complexity are also able to evaluate, reflect on and monitor the outcomes of their own behaviours. Higher levels of complexity enable individuals to conceive and imagine (i.e. anticipate) possible outcomes of own possible future behaviours. Then individuals can make deliberate choices about own future behaviours, plan ahead and develop intentions (Tomasello 2014). The possible future outcomes that individuals can anticipate and imagine in their experiencings also function as motivators and guides of individuals' current and future behaviours that thereby become *actions* (Bandura 2006). With increasing complexity of psychical systems, individuals can construct appropriate action plans and motivate and regulate their execution (Searle 2003). Through these abilities, individuals can increasingly become *actors* of their own lives who are able to partially choose and influence their own life circumstances as well as the directions and courses of their own development (Bandura 2006).

Individuals with more complex psychical systems can also conceive of their own psychical phenomena in and of themselves and reflect (within limits) on the operations that they use to process experiencings and to construct meanings and knowledge (e.g. abstraction, inference; Bandura 2006). As agents, individuals can also reflect on the experiences that they have made in relation to themselves and their own functioning and they can integrate these experiences in their autobiographical memories, thus expanding their psychical worlds by *psychical representations about themselves*—their selves (Gillespie 2006). Every individual becomes unique through his or her own self-related memories. On the basis of the continuity that individuals perceive in their memories and from which they conceive and imagine their possible futures, individuals construct their own personal identities and they develop and pursue goals and plans for their lives. By imagining possible events and outcomes in the more distant future, individuals make sense of their lives (Bruner 1986; Harré 1983; McAdams 2001; Thomae 1988).

Psychical representations about psychical phenomena are representations of a very special kind because, for psychical phenomena, physical properties cannot be conceived. There are thus no physical counterparts toward which these psychical representations could be approximated. Moreover, the "non-physicality" of psychical phenomena does not offer any point of reference that the introquesting individual could use to reliably differentiate in his or her continuously flowing experiencings various kinds of experiencings (e.g. those often referred to as thoughts, emotions, feelings, impulses) and various kinds of memorised resultants that he or she can reconstruct from his or her psychical system (e.g. those often referred to as self-concepts, attitudes, abilities, motives, interests, knowledge; cf. Kant 1786).

However, the specific formations that psychical representations take in any given individual are not important. What is important is the functionality that psychical representations have for the given individual in his or her given internal and external multi-contextual embeddedness. This functionality of psychical representations—their *meaning*—arises from the individuals' abilities to self-organise and to preserve themselves (Uher 2015a).

Socially Shared Psychical Representations

As members of the same species, individuals share much of their organismal properties—their basic physical systems (i.e. morphology, physiology, behaviour) as well as their general relations to the external physical phenomena that commonly occur in their particular habitat (i.e. their ecological adaptation; Uher 2011b). On the basis of their conspecifically shared organismal properties, individuals develop psychical representations that, despite their inherent idiosyncratic formation and development, show some functionalities that are generally similar. Individuals tend to conceive similar—i.e. conspecifically and thus socially shared—meanings that are functional for all members of their species (cf. Merleau-Ponty 1967; von Uexküll 1909). It is because of these similarities that individuals of humans and of other social animals are able to acquire knowledge—i.e. *learn*—from each other.

The most elementary forms of social learning presuppose that multiple individuals can directly perceive *one and the same* event of a given physical phenomenon, thus they presuppose phenomena that are extroquestively accessible. These forms of social learning also presuppose that individuals are able to externalise the *meanings* that they have constructed for their psychical representations and to produce externalisations (e.g. behaviours) that allow other individuals to correctly infer the individually constructed meanings. These other individuals must be able both to make the correct inferences and to construct similar meanings—each individual for him- or herself and in their own idiosyncratic formations. For example, from the temporal and spatial proximity between sight of a snake and an individual's warning call (i.e. externalisation; both extroquestively accessible), and on the basis of the already shared meaning of the call (e.g. "danger on ground"), other individuals nearby may infer that the caller may associate a snake with a danger and they may,

on the basis of this inference, construct pertinent associations in their own psychical systems. Predator-specific alarm calls are known from many species, such as rhesus macaques and vervet monkeys (Cook and Mineka 1989; Seyfarth et al. 1980).

This form of *social co-construction of psychical representations* occurs *in reaction to* particular events (e.g. snakes appearing). In addition, it depends on specific *constellations of occasions*. First of all, it depends on the temporal and spatial proximity of the relevant event and multiple individuals who can perceive this event and for which some individuals (the knowers) already construct a particular meaning, whereas others (the learners) do not. It furthermore depends on the learners' correct interpretation of the knowers' externalisations and of these externalisations' reference to the particular physical object being perceived in proximity as well as of the learners' perceived relevance that both the knowers' externalisations and the objects perceived have for themselves. The TPS-Paradigm therefore refers to this form of social learning as *reactive and occasion-based co-construction* of psychical representations. Phylogenetically seen, this may be the oldest form of how individuals co-construct psychical representations.

The abilities to become aware of and sensitive to others' perception and focus of attention (Call and Tomasello 2007) and to achieve joint attention to one and the same event (Tomasello 2009) further increase individuals' abilities to co-construct psychical representations. Individuals with psychical systems of higher complexity are also able to infer, from reflecting about themselves and from observing others' externalisations, that other individuals as well perceive and conceive of the world and develop psychical systems that may be analogous to their own (Gibson 1967; Schrödinger 1958). With increasingly complex psychical systems, individuals can infer others' intentions, take others' perspectives on the world and imagine themselves in the role of others, thus establish *intersubjectivity* (Mead 1934; Piaget 1928). Psychical representations about others' psychical systems, often referred to as *theories of mind* (cf. Whiten 1991), also enable individuals to substantially increase and refine their abilities to co-construct psychical representations.

13.5 Semiotic Representations: Tools Enabling Exchange Between Individual Minds

For co-constructing psychical representations (e.g. of external physical phenomena or of mental, emotional or volitional experiencings), individuals must overcome the unique intricacies that arise from the one-sided psyche-external surrounding connection and the fundamental imperceptibility of psychical phenomena by other individual's. Psychical phenomena, in and of themselves, as they are entirely internal and "non-physical", cannot directly interact with phenomena that are external to the individual's body—and thus not with other individuals' sensations and perceptions. To interact with external phenomena, individuals must *external*ise the meanings that they have constructed for their psychical events. That is, individuals must convert these meanings—through *external physicalisation*—into information in physical

phenomena that others can sense and perceive (i.e. access extroquestively), such as into information in the physical phenomena of behaviour (Uher 2013) and of matter other than those of the individuals' body (i.e. *objectivation*; cf. Moscovici 1961).

To externalise and communicate meanings that are essential for individuals' survival, species-specific behavioural repertoires that are shared by all conspecifics have evolved. These behaviours and the particular meanings that they convey are often acquired quickly and with little tolerance for error. Socially raised dog puppies learn quickly to correctly interpret snarls. In adult dogs, lack of this knowledge is rare, often resulting from social deprivation. Species-specific behaviours (e.g. snarls) typically refer to concrete physical events that are present in the given moments in which the behaviours occur (e.g. food, conspecifics) and they convey concrete meanings (e.g. defence readiness). Species-specific behaviours that are used to externalise vitally important meanings are *not completely arbitrary* but often linked with other behaviours of similar function and meaning. Snarls occur close to the teeth, which can be used to injure. Baring the teeth additionally supports the meaning conveyed by the snarling. Such links between functionally similar behaviours can reduce the risks of misinterpretation at least within a given species. The meanings that behaviours can convey are often species-specific (cf. the different meanings of physically similar behaviours of dogs and cats; Uher 2008a, b).

Importantly, the communication of meanings through species-specific behaviours presupposes *temporal and spatial proximity* of the individuals between which it occurs and of the physical objects to which the meanings refer. No dog snarls for defending food unless he perceives both food and a potential competitor. In addition, the possibilities to infer meanings are *bound to the* behaviours in which they are externalised—and thus *to the particular moments in which these behaviours occur*.

Genesis, Types and Metatheoretical Properties of Semiotic Representations

Meanings can also be externalised in external bodily activities or changes (e.g. vocalisations, movements) that have no a priori fixed (and likely evolutionarily derived) function in a given species and that thus need not be behaviours. Therefore, meanings can be assigned *arbitrarily* (cf. Holloway 1969) to such externalisations that thereby become functional—and thus behaviours. Through reactive co-construction, multiple individuals can psychically represent specific as*sign*ments in similar (i.e. socially shared) ways. The particular physical events (e.g. movements) that are used to externalise information from particular psychical events (e.g. constructed meanings) thereby become *signs* (e.g. semiotic behaviours).

Unlike species-specific behaviours, semiotic behaviours—i.e. *behavioural signs* (e.g. gestures, spoken language)—allow individuals to uncouple the transmission of meaning from the spatial and temporal coincidence of the particular physical events to which the meaning refers (e.g. snakes)—thus, from individuals' immedi-

ate perception of these events. This opens up further possibilities for individuals to co-construct socially shared meanings (see below). However, the transmission of meaning is still bound to the particular moments in which the behavioural signs are displayed. It is also still bound to the individuals who externalise the meanings (which, however, provides opportunities for other individuals to immediately check the inferences that they have drawn from the behavioural signs displayed).

Individuals can also externally physicalise meanings in matter other than that of their own bodies—i.e. in *material signs* (e.g. clothing, pictures, written language, numerals). Compared with behavioural signs, material signs are temporally more extended. As they are independent from individuals' bodies, material signs allow individuals to uncouple the processes of encoding meanings in signs from the processes of decoding the meanings from the signs again. This opens up unprecedented possibilities to *transmit meanings in absence of the objects* to which they refer and *across time and space* (though at the expense of possibilities to immediately check the inferred meanings with the individuals who have externalised the meanings). Material signs can therefore be conceived of as *physicalised resultants of past externalisations* of socially shared meanings.

Given that meanings are assigned arbitrarily to behavioural and material signs, the specific physical events used as signs have no immanent meanings in and of themselves. In behaviours, individuals' possibilities for creating sets of diverse yet distinct signs (e.g. phonemes) are constraint by their bodily abilities. For example, the phonemes of the words "dog", "chien", "cane" and "Hund", which denote the same animal in different languages[8], sound entirely different, but each of them centres on a vowel. By contrast, matter other than that of individuals' own bodies provides more opportunities for individuals to create sets of distinct signs (e.g. graphemes) because these materials are much more diverse and they can be designed and transformed. For example, across different spoken languages, humans have developed very different writing systems (e.g. Latin, Cyrillic, Greek, Arabic, Kanji or Hebrew alphabets). In some of these systems, single signs denote concrete phenomena or even associations of several concrete phenomena that denote abstracted phenomena. But in other systems, single signs denote only word stems, syllables or just letters that must be compounded into "words" before they are able to denote concrete and abstracted phenomena.

Behavioural signs and material signs can be created to externally physicalise *all kinds* of meanings. Communities en route to developing semiotic systems (most likely) first create signs that refer to concrete physical phenomena, the events of which can be directly perceived by multiple individuals (e.g. a dog, a lightning stroke). By referring to the physical phenomena as such, the meanings of such signs are *denotative* (i.e. literal). To both the signs and the physical phenomena that are being denoted by these signs, communities can assign additional meanings that are *connotative*. Connotative meanings often refer to the value that communities of individuals attribute to the phenomena to which the meanings refer or to their use (e.g. dogs as pets, guards or meat deliverers; cf. Bühler 1934; Shweder and Sullivan 1990).

[8] In English, French, Italian and German language.

Denotative meanings, as they refer to physical phenomena as such, are constructed on the basis of criteria that are bound to structures and associations occurring in nature and that therefore cannot be changed without introducing contradictions (cf. Bühler 1934; Shweder and Sullivan 1990). Horses are quadruped animals, bipedal animals are not horses. Material signs (e.g. drawings) denoting concrete physical phenomena (e.g. horses) are therefore not completely arbitrary because individuals can approximate their psychical representations to the physical properties of these phenomena that the individuals, given their shared organismal properties, likely perceive in similar ways. Therefore, denotative material signs often show similarities across sociocultural communities. Horses are typically depicted as quadruped creatures—even in Palaeolithic cave paintings up to 40,000 years old (Bahn 2007).

The construction of connotative meanings, by contrast, is not necessarily bound to structures and associations that can be found in nature. For some sociocultural communities, horses are meat deliverers, for others they are status symbols and for still other communities, horses are means for transportation or for carrying out heavy work. Variations in material signs (e.g. in horse paintings) may therefore result from and thus indicate variations in the connotative meanings that particular communities attribute to the phenomena that these signs denote. Variations in material signs can therefore be used to explore sociocultural differences in connotative meaning systems (as demonstrated, e.g. in analyses of children's drawings of their family members; Gernhardt et al. 2013). In language, however, such inferences are complicated because humans (most likely) first developed written language on the basis of behavioural signs (e.g. vocal and gestural language). The first material signs (e.g. written language) were developed only much later in human phylogenetic history and, most likely, these material signs were created to externally physicalise meanings for which behavioural signs had already been developed.

Over time, communities of individuals develop complex systems of systematically interlinked signs to which they assign denotative and connotative meanings. With increasing sophistication of the semiotic systems that have already been developed, communities are also able to establish signs that refer to phenomena that are not directly perceivable, such as properties that can only be inferred or abstracted from concrete events. Such communities can also create signs that refer to psychical phenomena in and of themselves, thus to phenomena for which physical properties cannot be conceived and that, moreover, can be perceived only by each individual him- or herself (i.e. introquestively). Given this, the pertinent material signs vary more strongly than signs representing concrete physical phenomena. The same applies to signs that refer to ideas of supernatural phenomena (e.g. spiritual beings) to which individuals ascribe properties that are incompatible with the properties of physical phenomena. Similarities in the pertinent signs developed by different communities (e.g. sculptures or pictures representing deities) can still be found because the social co-construction of inferred, abstracted and fictitious meanings presupposes signs that represent concrete meanings. Material signs representing spiritual beings therefore often show anthropomorphic or animalistic properties, thus properties of physical phenomena that can be directly perceived (e.g. physical beings).

To recapitulate, signs are created to represent meanings externally—thus necessarily physically (i.e. in behaviours or matter), which enables their direct perception by multiple individuals. Signs therefore facilitate the co-construction of meanings referring to concrete phenomena and first enable the co-construction of meanings referring to phenomena that cannot be directly perceived. Like all psychical phenomena, meanings are bound to the individuals who conceive them. Thus, although many signs are created in matters that are independent of individuals' bodies, semiotic representations are always *bound to the individuals who create and co-construct them*. The TPS-Paradigm therefore conceives of semiotic representations as *composite kinds of phenomena* that comprise psychical phenomena that are tightly intertwined with external physical phenomena that are used as signs. Hence, unlike behavioural or psychical phenomena in and of themselves, *semiotic representations are phenomena with heterogeneous metatheoretical properties* comprising both internal and external phenomena, both physical phenomena and "non-physical" ones and both phenomena that are bound to the present moment (e.g. spoken words, constructed meanings) and phenomena that are temporally more extended (e.g. written words, hieroglyphs).

Importantly, between phenomena with opposed metatheoretical properties, isomorphisms can generally be only low even if they are directly interrelated. Therefore, semiotic representations are phenomena with heterogeneous structures and of particular complexity. It follows that signs cannot be considered independently from the meanings that are assigned to them by particular communities, unless the physical phenomena used as signs are considered only as such rather than as signs (Uher 2015a, b, c).

Stability and Change of Semiotic Representations

Meanings, given their "non-physical" properties, can never be replicated in identical form as this is possible for the events of physical phenomena to some extent. Meanings must always be (re-)constructed anew by individuals in each given moment. Therefore, meanings *change* continuously.

Co-constructions of meanings and of semiotic representations are based on processes of ex*change* between two or more individuals—i.e. on *dialectical* transmissions of meanings. In dialectical processes, amongst others, interactions between elements can result in changes of the elements and their interrelations in and of themselves, thus leading to irreversible processes of change and development (Caprara 1996; Prigogine 1996). Processes of social ex*change*, given their dialectic properties, have their own dynamics that are unlikely to be isomorphic to the processes that occur in the psychical systems of each participating individual (cf. principle of emergence). Dialectical processes contribute to the particular dynamics that occur in the co-construction of socially shared meanings, thus fuelling their permanent and continuous change.

Importantly, changes in meanings need not correspond either to changes in the physical phenomena to which the meanings refer (e.g. dogs, horses, family members) or to changes in the behavioural or material signs that are used to externally physicalise these meanings (e.g. gestures, writing systems, sculptures). Meanings can also be forgotten, such as the meanings that ancient communities have once co-constructed and externally physicalised in material artefacts. Without sufficient knowledge about their creators, the meanings of many ancient artefacts cannot be reconstructed anymore and some remain mysterious for present-day humans. Meanings of socially shared representations can be only as stable as the mnemonic systems of the individuals who co-construct these meanings. Meanings decay with individuals' memories (e.g. in dementia) and with their lives. Communities therefore develop various ways to preserve and propagate their systems of shared meaning.

Communities of individuals can be conceived of as living systems. Communities organise themselves and develop structures and mechanisms enabling their members to systematically co-construct and propagate socially shared meanings and to preserve the functionality that these meanings have for them. For this purpose, human communities systematically install physicalisations in both individuals' behaviours (e.g. practices, institutions) and external matter (e.g. buildings; cf. World Installation Theory; Lahlou 2008, 2011). The possibilities to create, propagate and preserve semiotic systems depend on the particular kinds of external physicalisations used. Behavioural signs (e.g. gestures, practices, spoken language), as they are bound to the individuals—and thus always available (i.e. "at hand")—may be created more quickly than signs that are externally physicalised independently of individuals' bodies and for which additional materials are needed (e.g. stone, chisel, paper and ink). But behavioural signs, as they are bound to the present moment, may be more difficult to preserve and to propagate over longer distances. Variations in spoken language between regions and over time emerge more quickly than in written language and may therefore be more pronounced. Technologies to create material signs (e.g. stones inscriptions, letterpress, computers) increase communities' possibilities to create, propagate and preserve signs. Audiovisual technologies can also promote the propagation of behavioural signs—today even globally. When the behavioural sign of showing the soles of one's shoes to externalise anger and insult originating from the Arab world became known globally, it was soon also used by protesters in countries outside the Arab world and where it had previously been unknown, such as in Germany or the United States.

Social practices and institutions are mechanisms and structures that are aimed at governing individuals' behaviours in order to establish and to enforce social order and cooperation in communities (Durkheim 1895).

> Arbitrary symbols enforce consensus of perceptions, which not only allows [community] members to communicate about the same objects in terms of space and time…but it also makes it possible for social relationships to be standardized and manipulated through symbols. It means that idiosyncrasies are smoothed out and perceived within classes of behavior. By enforcing perceptual invariance, symbols also enforce social behavioral constancy, and enforcing social behavioral constancy is a prerequisite to differential task-role sectors in a differentiated social group adapting not only to the outside environment but to its own membership (Holloway 1969, p. 406; cf. also Baldwin 1896a, b).

However, the organising functions of social practices and institutions can be fulfilled and preserved only if all individual members of the community internalise (i.e. psychically represent) the meanings of the normative semiotic representations that are established for these purposes (e.g. norms, rules, rituals, religions) and only if the individuals coordinate their activities accordingly. As these normative semiotic representations are directed at individuals' *behaviours*, their meanings are primarily transmitted through behavioural signs. This requires individuals who physically represent the given institutions with their own bodies and behaviours, who communicate the normative meanings and who control and enforce individuals' adherence to these norms. Given the limitations of behavioural signs in terms of their boundedness to individuals and the present moment, larger communities and institutions also externally physicalise their normative semiotic representations in material signs (e.g. formal clothing, codices). But written norms and rules, in and of themselves, cannot affect anything unless individuals act upon them (cf. Weber 1922). This again reflects the heterogeneous metatheoretical properties of semiotic representations comprising both psychical and external physical phenomena (Uher 2015a).

Communities must also deal with the inevitable and continuous changes in meanings. If meaning systems that diverge too strongly from the physical phenomena to which they refer become established, and if meanings with non-proven, insufficient or even missing functionality are propagated and preserved, communities may become unable to self-organise and preserve their functional structures, both internally and across their boundaries so that, eventually, communities may collapse. The processes of creating, implementing and propagating social practises, norms, rules and institutions are important means for communities to self-organise. The temporal and spatial extensions of these processes have decelerating effects on the inevitable and continuous change of meanings. This gives communities the time needed to test the functionality of inventions and novel adaptations (see below; cf. Baldwin 1896a, b). Specifically, the processes involved in social exchange contribute to the communities' abilities to identify, promote and preserve those socially shared meanings that are important and functional for their survival (e.g. belief systems, religions), to adapt these meaning systems to changes occurring in their physical surroundings (e.g. economic or ecological changes) and to other (similarly changing) meaning systems within and beyond the boundaries of the particular community (e.g. social or ecological movements, political or societal systems) and to install the external physicalisations of their meanings systems accordingly (e.g. social practices, community buildings). These processes enable communities to influence and direct their own development, and thus to become actors in their own histories.

To preserve meanings, as they are not immanent to the physical phenomena that are used as signs in and of themselves, communities also develop semiotic representation systems in which meanings, especially normative ones, are encoded in contextualised ways, such as narrative histories of communities and nations (e.g. myths, legends) and their external physicalisation in textual documents (e.g. the Bible, the Koran, the Torah, national law codes). Contextualisation of physical representations can reduce but not prevent variations in the meanings that individuals

reconstruct from them; one and the same text is interpreted differently by different communities, leading to different behavioural practices. Supreme courts not only create new laws and physicalise them in textual documents; they also survey and decide on the interpretation of laws that are already textualised.

Significance of Semiotic Representations for Human Development

By enabling the transmission of experiences and knowledge between individuals and across times and places—though always bound to individuals' memories—semiotic representations open up new dimensions for development that increase individuals' opportunities to actively influence and create the conditions of their own lives.

With the creation of semiotic representations, individuals' development is no longer confined to their physical organismal properties that are derived in comparably fixed ways from genetically inherited molecular codes (Schrödinger 1944; Watson and Crick 1953). These genetic codes are generated from mechanisms of *random* variation and are acted on by selective external pressures (Darwin 1859; Wallace 1858). Genetic codes are also acted on by epigenetic processes enabling the transgenerational transmission of *physical* changes that individuals' have acquired over the courses of their lives (Mayr 1966; Waddington 1942; Bradbury 2003).

The organisational structures and functional processes of individuals' psychical systems, which are derived from each individual's interactions with and adaptations to his or her particular internal and external physical conditions, contribute to individuals' survival additional functionalities that are unequalled by any of the functionalities derived from their physical properties. These functionalities of psychical systems together with the external physicalisations of their central elements (i.e. their meanings) in behavioural and semiotic systems open up additional dimensions that affect individuals' development in highly complex ways and far beyond mere randomness and the physicality of matter (Bandura 2006; Jablonka and Lamb 2005). The complex composition of semiotic representations linking physical and psychical phenomena with one another accelerates and diversifies the processes of ontogenetic and phylogenetic development to an unprecedented extent, both quantitatively and especially qualitatively.

Most central to these additional functionalities is the individuals' organismal ability to self-organise their psychical systems—i.e. their *individual learning*. Baldwin (1896a) assumed that not the specific functionalities that individuals develop during ontogeny, in and of themselves, are selected on but rather individuals' general capacity for developing specific functionalities (cf. also Dobzhansky 1972).

Of particular significance are the individuals' capacities for developing psychical processes that are conscious. Conscious psychical processes go beyond the psychical properties that individuals develop on the basis of congenital properties, enabling individuals to survive on the basis of spontaneous activities (e.g. instincts; Baldwin 1896a). The psychical properties that individuals develop from their interactions with the particular internal and external physical conditions that they encounter in

their lives contribute to individuals' survival—and thus also to the maintenance of those heritable organic variations that enable the development of these psychical properties. "This principle secures by survival certain lines of *determinate phylogenetic variation* in the directions of the *determinate ontogenetic adaptations* of the earlier generation" (Baldwin 1896a, p. 447; emphases added), thus preventing "incidences of natural selection" and allowing for more time for the population to produce variations, both novel and congenital ones (Baldwin 1896a, b).

Unless the organic variations that enable the individuals of a given population to develop particular kinds of psychical properties have already become genetically heritable (e.g. properties of sensual perceptions), individually developed psychical properties are functionalities of novel kind (e.g. mathematical abilities). Thus, it is by using these novel functionalities that the individual can survive and propagate the novel kinds of psychical properties in his or her population, such as by externally physicalising the meanings of these functionalities in behavioural or material symbols (e.g. mathematical symbols; cf. Peirce 1902). External physicalisations allow individuals to build on these novel functionalities in order to develop additional and more complex psychical properties with even more complex functionalities (e.g. architectural and engineering abilities). In this way, individually developed psychical properties with *proven functionality for the particular internal and external physical conditions that are present in individuals' lives* rather than only random mutations of a priori neutral (i.e. blind) functional value can become gradually and transgenerationally incorporated into the genetically and epigenetically transmissible organic variations of the population, while the principle of natural selection may be still operative (Baldwin 1896a; Tomasello 1999).

In socially living species, individual development into *determinate* directions rather than into random ones is also enabled by the "purely extra-organic ways of social heredity" (Baldwin 1896a, p. 539)—i.e. by transmissions through behaviours and semiotic representations (Jablonka and Lamb 2005). Both organic and non-organic inheritance may contribute to the *same psychical functionality*; individuals' conscious psychical processes may lead them to consciously do what they may also do congenitally (i.e. instinctively). Transgenerational transmissions through behaviours and semiotic representations also enable populations to preserve functionalities that either are not yet or never will be organically heritable. Although this kind of transmission is not based on organic heredity in and of itself, it keeps alive heritable organic variations. It "thus sets the direction of ontogenetic adaptation, thereby influences the direction of the available congenital variations of the next generation, and so determines phylogenetic development" (Baldwin 1896a, p. 537).

Of particular significance are individuals' psychical abilities to co-construct psychical representations because socially shared representations enable coordinated activity and cooperation between individuals (Lahlou 2001). Individuals who are able to develop novel psychical properties that promote social coordination and cooperation have therefore advantages over individuals who are less or (still) not able to develop such psychical properties. This creates new—i.e. social—selection pressures that determinate the directions for future development in a given population. In addition, these social selection pressures may also raise the functionality

of a social community to a new standard such as by enabling novel behavioural performances (e.g. metalworking), thus changing the frame of selection within and across communities and populations (Baldwin 1896a, b; Dennett 1991).

Within a given population, as behavioural and semiotic transmissions become more important for propagating the development of psychical properties that enable novel and significant functionalities, capacities for developing specific psychical properties on the basis of heritable organic variations (e.g. instincts) become more broken up to allow for the plasticity that individuals need for their individual learning. "The [human] child is the animal which inherits the smallest number of congenital co-ordinations, but he is the one that learns the greatest number" (Baldwin 1896a, p. 540).

Individual learning allows for the development of highly individualised psychical systems (often referred to as "personality"; Uher 2015a, b, c) that are idiosyncratically adapted to the particular physical conditions that the given individual encounters in life, both internally (i.e. the own body and its particular organic variants) and externally (e.g. ecological system, physical installations of the sociocultural community). This high plasticity in the individual development of adaptations during ontogeny, as compared with the adaptations enabled by instincts, is considered one of the driving forces behind humans' accelerated phylogenetic development. Through behavioural and semiotic externalisations, individually developed functionalities and knowledge can be passed on to other individuals, thus preventing social retrogression and enabling cultural evolution (Jablonka and Lamb 2005; Tomasello 1999).

For enabling such transmissions, young individuals must have the capacities to develop their psychical properties not only from processing their *own* experiencings and from constructing their *own* meanings but also from reconstructing and processing meanings that *other* individuals have behaviourally and semiotically encoded—through *observational learning, instruction and education* (Baldwin 1896c). This entails particular challenges for young individuals because these meanings were constructed by *other* individuals—thus originate from other psychical systems that are each highly individualised and historically unique. Young individuals must be able to develop the psychical abilities that are needed to infer and to reconstruct these meanings from others' behavioural and semiotic externalisations, to process and psychically represent these reconstructed meanings and to develop pertinent psychical properties that are sufficiently functional for themselves and in their own lives. In addition, young individuals must psychically represent the particular assignments of these socially shared meanings to the particular behavioural and material signs that are established in their community. This means, young individuals must organise the structures of their psychical systems not only in *egocentric* but also in *allocentric* ways.

Individuals with pertinent inherited properties can develop complex psychical systems through reactive co-construction. By observing their mothers and other individuals, young individuals can develop complex psychical properties, enabling them to acquire complex functionalities that have been developed by individuals of previous generations. In this way, individuals of various species of non-human primates, amongst others, learn to use tools to crack open oysters, crustaceans or nuts

(e.g. crab-eating macaques, Gumert et al. 2009; capuchin monkeys, Visalberghi and Fragaszy 2012 and chimpanzees, McGrew 1992).

In humans, however, the psychical functionalities that have been developed by previous generations (e.g. mathematical abilities, natural science knowledge) and the psychical abilities that are necessary to acquire and to successfully apply these psychical functionalities have meanwhile reached such levels of complexity that these functionalities cannot be transgenerationally transmitted by means of reactive co-construction alone. Their transmission additionally requires processes of co-construction in which still less capable individuals are exposed to purposefully induced (rather than incidental) co-occurrences of events and are actively guided in their psychical development (i.e. their learning) by individuals who already have developed the particular kinds of psychical functionalities and capabilities. This is referred to as *active and intention-based co-construction* in the TPS-Paradigm.

Active intention-based co-construction and guidance (i.e. instruction, education) presuppose the ability to make valid and differentiated inferences from the externalisations (e.g. behavioural performances) of other individuals—especially of (still) less capable ones—on the hitherto reached psychical properties of these individuals (cf. theories of mind; Baldwin 1906). More capable individuals can actively guide the learning of (still) less capable individuals by making such inferences from the learners' previous externalisations, by considering possible constraints and opportunities that are available to the learners in the given settings and moments and by mentally constructing the learners' potential capacity for developing particular kinds of psychical properties in the near future (cf. zone of proximal development; Valsiner 1987; Vygotsky 1978). With increasing capabilities and levels of complexity of their psychical systems, young individuals are increasingly able on their own to further differentiate their psychical properties and to develop novel functionalities by encoding behaviourally and semiotically encoded meanings (e.g. by studying textbooks).

Semiotic systems comprise both *physicalised resultants of past externalisations* of socially shared meanings (i.e. signs) and co-constructed memorised psychical resultants derived from past individual experiencings (i.e. socially shared meanings). As resultants of past lives, semiotic systems implicitly reflect the experiences and the knowledge made and created by individuals of previous generations who were exposed to other internal and external physical conditions and who have lived in other times (cf. Gergen 1973; Peirce 1902; Valsiner 2012; Vygotsky 1934). Thus, the meanings of signs are derived from the past—even the meanings of the signs that are used to refer to the present and even though their particular meanings will inevitably have changed over time within and across communities.

In a nutshell, semiotic representations are of crucial significance for individual development and human evolution. Semiotic systems contribute to human's ability to actively create external physical conditions that are highly complex and rapidly changing and that, in turn, require human individuals to continuously develop and propagate novel psychical functionalities. As these processes are inherently irreversible and historically unique, they increase and accelerate processes of diversification in the developmental history of life.

13.6 Present Thyself: Artificially Modified Outer Appearances—Semiotic Representations of Special Kind

An obvious core characteristic of the human species found in all communities around the globe is the artificial modification of individuals' natural external morphology (e.g. hairstyle, body painting, fragrances, clothing, accessories). The TPS-Paradigm refers to these modifications as the phenomena of individuals' artificially modified outer appearances. These phenomena are external to individuals' bodies and primarily material physical. They are used to change individuals' outer appearances selectively on an individual level and in addition to those changes that emerge naturally during ontogenetic development (e.g. in body size, shape and proportions; cf. Uher 2013, 2015a). Besides some functions for protection and warmth, these phenomena are often used to convey particular meanings to other individuals. In fact, they are often targeted toward others' perceptions, thus playing important roles in social perception. The TPS-Paradigm therefore conceives of artificial outer-appearance modifications as special kinds of semiotic representations comprising both psychical phenomena (i.e. meanings) and external material phenomena that are attached to individuals' bodies (e.g. clothes), in contrast with other material signs.

The temporal extension of the phenomena of artificially modified outer appearances differs from that of other external physical phenomena of individuals' bodies. Specifically, individuals can artificially modify their outer appearances far more quickly than natural changes can occur in their external morphology. But compared with the fluctuating and momentary phenomena of behaviours, artificially modified outer appearances are much more temporally extended, which facilitates their perception by others, thus promoting their semiotic function. Individuals use artificial outer-appearance modifications to physicalise meanings of normative semiotic representations, such as to indicate their membership to a particular social community or their social status within the given community (e.g. uniforms, insignia). As these phenomena can be modified by the individual him- or herself, artificial outer-appearance modifications are also used to construct meanings that have particular relevance for their carrier, such as to (co-)construct the individual's "personality"—both by him- or herself and by others (Uher 2015a, b, c).

13.7 Contexts: "Environments" That Are Inseparable from Individuals

The phenomena of contexts, in the broadest sense, refer to the events that are considered with regard to particular phenomena in a given individual and that are commonly referred to as "environments", surroundings, circumstances, conditions, background or settings, amongst others. In everyday life—but also in science—individuals often conceive of the "environment" as the external physical events (Gifford 1997) that surround or encircle (i.e. environ) the individual being consid-

ered, thus conceptually separating the given individual as the actual target of consideration from all other phenomena (Valsiner 1987). This dualistic concept likely reflects human individuals' experiences of themselves as agents who encounter, actively seek out and also create their conditions of their lives. Individuals always perceive these conditions from the particular viewpoints that are enabled by their own psychical systems. Given the particular perceptual and conceptual representations that individuals have developed of their world, they also tend to conceptually separate specific parts of the "environment" from one another, such as abiotic from biotic parts (i.e. non-living versus living matter) or natural from social and built parts (i.e. nature versus culture). But such differentiations always depend on the particular conceptual perspective taken by a given individual in a given situation. Specifically, the biophysical "environment" also comprises non-conspecific and conspecific—thus social—settings (e.g. animal and human family members are present not only with their bodies but also with their psychical systems and social relationships). Vice versa, the socioculturally built "environment" also comprises abiotic and biotic physical phenomena (e.g. cities, gardens, agriculture).

Exclusive conceptual separations of individuals from their "environment", although they may appear obvious from each individual's own viewpoint, are not tenable, however, because the same external physical condition is not the same for all individuals (Lewin 1936). Von Uexküll (1909) therefore differentiated *Umgebung*, the given-around or surrounding, from *Umwelt*, the world-around. The *given*-around or surrounding is conceived of as the physically given in which organisms are included as physical objects. The *world*-around, by contrast, is determined by the individuals being considered and their particular organismal properties that enable them to perceive only particular properties of their surroundings. Hence, the world-around is not just physical and not just external to the individual. As self-organising living organisms, individuals are so intimately interconnected with the external physical surrounding that both cannot be conceived of independently from one another. As a consequence, elements of the external physical universe cannot be conceived of as being *exclusively separated* from the individuals who are being considered. Instead, they can only be conceived of as *inclusively separated* (Valsiner 1987).

Inclusive conceptual separations can be made on the basis of the particular forms that can be conceived with regard to the three metatheoretical properties for the different kinds of phenomena that are involved in an individual's world-around (Uher 2015a, c). The TPS-Paradigm therefore conceives of the *phenomena of contexts as composite kinds of phenomena* that comprise at least one basic kind of phenomenon (i.e. morphological, physiological, behavioural or psychical), which is thus physically inseparable from the studied individual's body (without destroying its integrity). In addition, a given contextual phenomenon may comprise further basic kinds of phenomena and/or external phenomena that are independent from the studied individual's body (e.g. family members, books, interiors). Hence, contextual phenomena may comprise both physical and psychical phenomena, both external and internal phenomena as well as both phenomena that are bound to the immediate moment and phenomena that are temporally more extended. Given these heterogeneous metatheoretical properties, isomorphisms between interrelated events of the different kinds of phenomena that are comprised by contextual phenomena are necessarily low.

Concepts of Situations and of Their Behavioural and Psychical Relevance

The TPS-Paradigm conceives of an individual's situation as the particular constellation of the internal and external events that are present in a given moment and that the individual can therefore directly perceive. With regard to explorations of experiencings, this concept considers the fact that psychical events are dynamically interrelated to and co-determined by all concurrent physical events both internal and external to the individual as well as by the physical and psychical resultants that the individual has retained from past events in his or her physical and psychical systems. Specifically, as the individual's body is always present and interrelated with his or her psychical events, an individual's situation always comprises internal physical events and physical resultants (e.g. blood sugar level, nutritional condition, health status, etc.). The universe of all external physical events may be infinite, ranging from the immediate surrounding (e.g. microorganisms on the skin surface, food on the table) over the conditions that are present on this planet (e.g. climate, world population) up to events in the outer space (e.g. solar wind, comets). From this universe of external physical events, only those concurrent events form part of the individual's situation that are immediately present in a given moment such that the individual can, at least theoretically, directly perceive these events (whether consciously or not). Finally, from the universe of the individuals' memorised psychical resultants, a situation comprises only those elements that he or she retrieves and reconstructs in his or her experiencings in the given moment.

Similarly, in his field theory, Lewin (1936) conceived of "life-space" as the entirety of all internal and external influences on the individual in a given moment that dynamically interact with one another and that are governed by psychical forces. From this "life-space", Lewin distinguished the "foreign hull" that he conceived of as all those external physical influences that are not governed by the individual's psychical properties (Lewin 1936, p. 73) and that are thus not perceived by or not relevant for the individual.

Behavioural situations[9] are specified in the TPS-Paradigm as a particular kind of situation that is conceived of as the constellation of those particular external physical events that functionally mediate the individual's external changes or activities in a given moment—i.e. his or her behaviours. Thus, behavioural situations, in and of themselves, are external to the individual. But the criterion for demarcating from the universe of all external physical phenomena those particular events that constitute a behavioural situation for a given individual in a given moment is bound to properties of that individual. This criterion is defined as the effectiveness with which external physical events make functional the individual's external changes and activities that thereby become behaviours. Thus, in the concept of behavioural situations, external physical events are *separated only inclusively* from the individual and these inclusive conceptual separations are made on the basis of the different metatheoretical

[9] Previously called the "environmental situation" (cf. Uher 2013; Uher et al. 2013a).

properties that can be conceived for the different kinds of phenomena that behavioural situations comprise.

Importantly, this demarcation is based on individuals' bodily events (i.e. on behaviours) rather than on internal physical or psychical events. As both the individuals' behavioural events and the physical events in their external surrounding can be directly perceived by multiple individuals, these events can be directly related to one another for exploring which particular external physical events are, in fact, functionally mediating particular behavioural events in a given individual and moment. This is important, as the specific events that constitute a behavioural situation need not be consciously perceived or be known a priori either by the individuals being considered or by those individuals observing them (e.g. parents, scientists).

Given that individuals' psychical systems can be conceived of as "non-physical", self-organising and in parts also self-referential, the diversity of the psychical properties that individuals may develop and that they can also combine with one another may exceed by far the diversity of the behavioural properties that individuals can produce, given the physical constraints of their bodily abilities. Individuals may therefore develop more diverse interrelations between external physical events and their psychical properties than they may be able to establish between external physical events and their behavioural properties. Hence, as with semiotic representations, it is the psychical phenomena that constitute the essential component of situations. But given the fundamental imperceptibility of psychical phenomena by other individuals, exploring and understanding the ways in which behavioural events enable individuals to bridge the one-sided psyche-external surrounding connection is essential for exploring individuals' psychical systems (cf. Uher 2013, 2015a, b, c).

Behavioural situations, as they mediate the individuals' behaviour in a given moment, are always also psychically relevant to them. But conversely, not every situation is also behaviourally relevant for individuals. Reading a book hardly involves any behavioural events at all (e.g. turning a page, saccadic eye movements). But through perception, (capable) individuals can semiotically encode from the particular external physical events that are present in this situation (e.g. printed words in a book) a lot of new information that are psychically relevant for them in the given moment and that enable these individuals to further develop their psychical properties and to construct new knowledge. The individuals interact with the external physical events only on the basis of their particular physical properties (e.g. sensory ones) and their particular psychical properties (e.g. perceptual and conceptual ones), thus only *internally*. Such internal interactions enable individuals whose psychical systems have already reached a certain degree of complexity to further develop their psychical systems without necessarily having to behaviourally, thus externally, interact with these events as well (e.g. physically disabled or paralysed individuals can read and learn from books and can also make science such as Stephen Hawking). In most situations, however, individuals interact with external physical events both internally and externally yet to a varying extent.

The ability to internally interact with external physical events entails particular intricacies for investigations of the psychical phenomena of individuals who suffer from complete loss of voluntary motor control and who therefore cannot externally interact with their external surrounding (e.g. patients suffering from locked-in syndrome; Laureys et al. 2005). The inability of these individuals to produce behaviours considerably complicates the physicians' possibilities to correctly diagnose these pathological conditions—i.e. to make the correct inferences on these individuals' psychical events. This illustrates the outstanding importance and the direct relevance that explorations of individuals' behaviours have not only for explorations of their psychical phenomena in exceptional circumstances but also in everyday life situations (Uher 2013, 2015a).

The situational concepts provided by the TPS-Paradigm differ from some previous concepts. For example, the concept of "psychological situations" (Rotter 1954, 1981; Shoda et al. 1994) denotes external physical properties that are subjectively relevant to the individual yet without differentiating the particular involvement of the individuals' behavioural and psychical events. But in line with the concepts of this paradigm, the concept of "psychological situations" is built on the recognition that particular external physical events have, in different constellations, different relevance for different individuals. The TPS-Paradigm complements this recognition by showing that the relevance that situations have for individuals' behaviours differs from the relevance they have for individuals' psychical systems.

The particular constellation of external physical events that are psychically and behaviourally relevant to a given individual in a given moment (i.e. that constitute the individual's situation), in and of themselves, can but need not be bound to the present moment. These events become part of the individual's situation only *while* the individual is internally and externally interacting with these events—i.e. *while* the individual is perceiving these events and *while* they are mediating the individual's behaviour. Before and thereafter, these events conceptually belong to the universe of external physical events. This immediacy corresponds to Lewin's (1936) concept of "life-space" and the actuality of its functioning.

Events in the individuals' wider contextual layers of that external universe (e.g. *socioeconomic, sociocultural or societal systems*; cf. Bronfenbrenner 1979) can affect individuals only indirectly as mediated through the physical events that are present in a given situation so that the individuals can directly perceive them (e.g. goods in the supermarket, governmental decisions publicised in the print media, the behaviour and body of institutional representatives, such as policemen). Once individuals have psychically represented the meanings of such events happening in their wider surrounding, these events can also further affect the individuals (again only indirectly) when the individuals revive their pertinent psychical representations in a given moment and situation (e.g. when individuals recall what they know about the ways in which the goods that they find in the supermarket have been produced).

Likewise, individuals themselves can form part of each other's situations only if they can directly perceive others' externalisations or revive others' past externalisa-

tions from their own memories (e.g. remembering someone's request to buy a particular product in the supermarket). Today, communication technologies (e.g. phone calls, emails, Skype) enable individuals to transmit their externalisations over large distances all around the globe so that individuals' immediate bodily presence is no longer required for transmitting externalisations, as has been the case for most of human evolutionary history.

Depending on individuals' organismal properties, particular external physical events enable individuals to perform particular activities that are called *affordances* (Gibson 1979). One and the same external physical setting (e.g. a text book) affords very different possibilities for activities to individuals of different age groups (Gibson 1979) and different species (von Uexküll 1980), for both solitary and social activities (Gaver 1999). Moreover, individuals themselves can offer for one another possibilities for activity through their bodies and their externalisations (e.g. soccer games, choirs; Valenti and Good 1991). The TPS-Paradigm therefore conceives of affordances as *potential behavioural situations*. The entirety of those affordances that are actually perceived by a given individual and involved in his or her behaviours in a given moment constitutes the behavioural situation of that individual.

Present-day humans also actively create in their external surroundings affordances that are aimed at matching individuals' particular psychical properties and at promoting their self-preservation and prosperity, such as external physicalisations of semiotic representations (e.g. text books) and other physical installations (e.g. school rooms, socio-technical systems in agriculture and industry; Lahlou 2008, 2011). These systems in turn, influence, prestructure and organise individuals' lives. Thus, present-day human individuals live in conditions largely of their own making (Bandura 2006).

Individuals' perceptions of their external physical surroundings, the affordances both naturally present and artificially created as well as individuals' psychical representations of these affordances continuously change and develop over time. Therefore, contextual phenomena are not only species-specific (von Uexküll 1909) and individual-specific (Rotter 1954, 1981; Gibson 1967) but also culture-specific (Barker 1968; Hall 1966).

Cultures: Systems of Semiotic Representations

Cultures, in the broadest sense, denote semiotically mediated systems of socially shared meanings (Geertz 1973; Weber 1904), thus systems of semiotic representations. The TPS-Paradigm therefore conceives of cultures as *composite kinds of phenomena* comprising psychical phenomena (i.e. individuals' psychical representations of socially shared meanings) and external physical phenomena (i.e. behavioural and material signs) in which these meanings are externally physicalised (e.g. behavioural practices, written documents, sculptures, monuments). In each of these different kinds of phenomena, some events are bound to the present moment, whereas others are not. As meanings are bound to individuals' psychical systems but

not immanent to the signs in which they are externally represented, individuals must always reconstruct anew in their ongoing experiencings the particular meanings that particular signs have for them. Therefore, and given the dynamic processes of social exchange in which cultural representations are being developed, cultural meanings are continuously changing—despite their physicalisation in external matter. Like all semiotic representations, cultures are therefore phenomena with heterogeneous metatheoretical properties and highly complex structures in which isomorphisms between interrelated events of the different kinds of phenomena that they comprise are low.

Hence, *cultural phenomena always involve psychical* phenomena and therefore cannot be conceived of as exclusively separated from the individuals who create and use them (Valsiner 1987). Language, one of the most important kinds of human cultural phenomena, is conceptually inseparable from the conscious workings of human individuals' psychical systems (cf. Geertz 1973). Culture can be described as

> …an organizing principle of each and every human mind, in any society. It is thus everywhere—always in action, but usually rarely noticed. We do not notice the most basic and ordinary facets of living (Valsiner and Han 2008, p. 3).

Cultural representations, as they comprise both psychical and external physical phenomena, allow for transmitting meanings across individuals, time and space. This enables cultural representations to contribute to the directed individual development within and across generations—i.e. to the "cultivation of individuals through the agency of external forms which have been objectified in the course of history" (Levine 1971, p. xix, citing Simmel). The social exchange processes that are involved in the creation, maintenance and propagation of cultural representations have important functions for the communities' self-organisation and self-preservation. These processes enable communities to canalise the continuous processes of change that are inherent to both the individual and the socially shared construction of meaning, such as by creating a common "cultural identity" (cf. Geertz 1973).

The reiterative processes of encoding and decoding meanings that are necessary for developing cultural representations involve repeated conversions of information between psychical phenomena (i.e. meanings) and their external physicalisations (i.e. in behavioural and material signs; cf. Uher 2015a, c). These processes also promote the creation of novel meanings, thus novel psychical functionalities.

> In cultural formations, the mind has reached an objectivity that makes it independent from the coincidences of subjective reproduction and, at the same time, subservient to the central purpose of subjective accomplishment[10] (Simmel 1919, pp. 223–253).

Writing is a means not only to externally physicalise psychical representations that individuals have already developed but also to develop new psychical properties and to create new knowledge, thus novel meanings and functionalities. Writing

[10] Original: "In den Kulturgebilden hat der Geist eine Objektivität erlangt, die ihn von allem Zufall subjektiver Reproduktion unabhängig und zugleich dem zentralen Zweck subjektiver Vollendung dienstbar macht."

therefore is an inherent part of scientific work. But the physicalised end product is not the essence of cultural (i.e. semiotic) representations. The driving forces propelling individual and community development arise from the continued and iterative processes of back and forth conversions of information between meanings constructed by individuals and their external representation in physical phenomena that other individuals can perceive and from which these other individuals can reconstruct, integrate and further develop these meanings in their own minds.

> At the end of the pathway, there is not the artefact, in the persistent existence of which the creative process is frozen, but there is the "you", the other subject that receives the artefact for including it into his or her own life and thereby transforming it back into the medium from where it originates[11] (Cassirer 1942/2011, p. 114).

13.8 Summary and Conclusions

This article applied the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals (TPS-Paradigm) to explore different kinds of phenomena and their perceptibility by individuals in everyday life. A particular focus was placed on exploring the properties that make psychical phenomena unique among the phenomena of life.

The Unique Properties of the Phenomena of the Psyche

Like all phenomena of life, psychical phenomena occur naturally in individuals of all age groups and across species. But unlike all other phenomena of life, their immaterial properties cannot be conceived of as being physical (i.e. featuring spatial units or at least rather constant interrelations to the physical phenomena to which they are connected) and that are therefore referred to as "non-physical" in the TPS-Paradigm.

Also unlike all other phenomena of life, psychical phenomena cannot be directly perceived by multiple individuals (i.e. extroquestively accessed) as is possible for all physical phenomena. Psychical phenomena are perceptible only by each individual him- or herself through introquestion; in other individuals, they can only be inferred from individuals' externalisations. But externalisations are phenomena of different kind for which different metatheoretical properties can be conceived and that are therefore unlikely to be isomorphic to the psychical phenomena to which they are related. This precludes straightforward inferences from individuals' externalisations to their psychical phenomena, especially if in-

[11] Original: "Denn am Ende dieses Weges steht nicht das Werk, in dessen beharrender Existenz der schöpferische Prozess erstarrt, sondern das "Du", das andere Subjekt, das dieses Werk empfängt, um es in sein eigenes Leben einzubeziehen und es damit wieder in das Medium zurückzuverwandeln, dem es ursprünglich entstammt."

dividuals are concerned who cannot report about themselves (e.g. young children, animals) and who thus cannot validate the inferences made by others. It follows that psychical systems cannot be studied without also exploring individuals' externalisations—their behaviours and (in humans) their semiotic representations.

Individuals' access to their own psychical phenomena—both ongoing experiencings and memorised psychical resultants (i.e. experiences)—is strictly bound to the present moment, the here and now. Experiencings are highly ephemeral and fluctuating. There never is a moment to pause or at least to slow down the continuous and irreversible flow of events to become more fully aware and to reflect on the events experienced. At the same time, and awareness and reflection inevitably introduce changes to the course of experiencings. These peculiarities preclude any possibilities for systematic and undisturbed self-explorations (Kant 1786; Wundt 1904). Moreover, experiencings are being continuously processed and integrated into the individual's psychical system that thereby continuously changes and develops in self-organising and partially self-referential ways and largely uncoupled from the physical laws to which physical phenomena are bound. The systematic structures and functionings emerging from these processes are therefore intrinsically idiosyncratic.

Psychical abilities enable individuals to internally represent the physical phenomena encountered in life also in the moments and situations in which these phenomena are not present, thus disembodied from their immediate perception. Therefore, and because of the "non-physical" properties that can be conceived for psychical representations, individuals can psychically operate and transform the represented information and properties in ways that are not enabled by the physical phenomena that are being represented in and of themselves. Individuals can therefore make inferences to properties that are not directly perceptible in the given physical phenomena. Individuals can make abstractions, comparisons and generalisations to develop conceptual representations of abstract properties that need not have direct counterparts to concrete physical phenomena that can be directly perceived. From processing psychical representations, new structures and qualities can emerge, enabling new functionalities for individuals' abilities to organise and preserve themselves in the particular internal and external contexts of their own lives—and already within their own lifetimes.

But these peculiarities also entail intricate challenges for individuals' abilities to communicate and exchange about their psychical properties. Different species and, in particular, humans have evolved various solutions to overcome these challenges, such as passive occasion-based and active intention-based co-constructions of psychical representations. The creation of behavioural and material signs to externally physicalise individually constructed meanings enabled humans to systematically communicate about their psychical properties—despite the imperceptibility of these properties by other individuals—and, in doing so, to propagate inventions across time, places and generations. These unprecedented possibilities opened up new pathways, enabling humans to actively select, influence and create the conditions of their own development—as individuals, communities and species.

Transdisciplinary and Philosophy-of-Science Perspectives: Opening Up New Avenues for Exploration

The unique constellation of properties characterising psychical phenomena entails challenges unknown to other sciences. Even more so as it is precisely these phenomena—and no others—through which all science is made.

The TPS-Paradigm applied in this research explicitly considers the fact that it is human individuals who make science and that thus, any scientific endeavour inherently depends on and reflects the particular psychical abilities of human individuals—and their limitations. Among all the many fallacies, biases and errors in reasoning that are intrinsic to human minds and known so far, the TPS-Paradigm considers those that are particularly relevant for scientists exploring individuals and their psychical systems.

To limit the biases derived from the individual scientists' own particular perspectives on their objects of research, given their own background in particular scientific disciplines and research traditions, the TPS-Paradigm adopts transdisciplinary perspectives to develop comprehensive research frameworks that can be used and refined by scientists across disciplines. To first enable comparisons between concepts and methodologies from different disciplines in order to identify commonalities and differences and to enable coherent integrations into interrelated frameworks, the TPS-Paradigm adopts philosophy-of-science perspectives to scrutinise the very means by which scientific knowledge is generated. It aims to make explicit as comprehensively as possible and to scrutinise the absolute presuppositions that scientists make about their objects of research (e.g. about individuals as living organisms such as the principle of emergence) and the metatheories and methodologies that scientists use to gain knowledge about these objects (e.g. the epistemological principle of complementarity). The interrelated philosophical, metatheoretical and methodological frameworks provided by the TPS-Paradigm are targeted at supporting scientists to critically reflect, discuss and further develop previously established theories, models and research practices and to derive strategies and ideas for novel lines of research in the future.

With the consideration of three central metatheoretical properties—(1) location in reference to the individual, (2) temporal extension and (3) physicality versus "non-physicality"—that can be conceived in different forms for all phenomena and that determine the phenomena's perceptibility by individuals, the TPS-Paradigm provides an elementary system that can be used straightforwardly to differentiate various kinds of phenomena from one another on a metatheoretical level. On the basis of this system, the TPS-Paradigm differentiates the phenomena of morphology, physiology, behaviour and the psyche. These phenomena are conceived of as basic kinds of phenomena because they are physically inseparable from the body of the intact individual and because each given kind of phenomenon has rather homogeneous constellations of metatheoretical properties. It was shown that, given their different perceptibility by individuals, some of these phenomena should be differentiated from one another more clearly than previously done, such as psy-

chical from behavioural phenomena and experiencings from memorised psychical resultants (as two kinds of psychical phenomena), in both theoretical and especially empirical research.

It was also shown that, conversely, conceptions of some other kinds of phenomena as being mutually exclusive and distinct are not warranted, such as the widespread dualistic conceptions of person versus situation, individual versus "environment" and "personality" versus culture—thus, of nature versus nurture (Uher 2015c). To conceive of such phenomena, the TPS-Paradigm introduces the concept of composite kinds of phenomena that comprise several different kinds of phenomena, among them at least one basic kind of phenomenon, and that may also comprise external physical phenomena that are independent of the studied individuals' bodies (e.g. material signs). The phenomena of semiotic representations (e.g. language), artificially modified outer appearance (e.g. clothing, hairstyle) and contexts (e.g.situations, "environment") are conceived as composite kinds of phenomena. This metatheoretical concept explicitly considers that particular kinds of phenomena are tightly interrelated and that from these tight interrelations new properties emerge (e.g. semiotic properties, abstractions). These new properties cannot be understood if the different kinds of phenomena that are involved are considered independently from one another. This novel concept allows to explore these interrelations by enabling inclusive conceptual separations that are made on the basis of the particular metatheoretical properties that can be conceived for each given kind of phenomenon of which composite kinds of phenomena are composed.

Building on the philosophical and metatheoretical elaborations presented in this research, a subsequent article (Uher 2016) will derive methodological implications for scientists exploring individuals and the workings of their minds. It will elaborate basic methodological principles that meet the particular challenges identified to comprehensively investigate each given kind of phenomenon explored in individuals, again putting the main focus on psychical phenomena. The article will present and illustrate methods from various fields of research that are suited for the empirical implementation of the methodological principles elaborated and it will outline ways in which suitable methods that have yet to be developed could be devised in future research.

Acknowledgments The author acknowledges funding from the Deutsche Forschungsgemeinschaft DFG (UH249/1-1).

References

Adam, M., & Hanna, P. (2012). Your past is not their present: Time, the other, and ethnocentrism in cross-cultural personality psychology. *Theory and Psychology, 22,* 436–451.

Adams, D. K., & Zener, K. E. (1935). Translators' preface. In K. Lewin (Ed.), *A dynamic theory of personality. Selected papers*. New York: McGraw Hill.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45,* 1363–1377.

Ader, R. (2006) *Psychoneuroimmunology* (Vol. 1, 4th ed.). New York: Elsevier.

Althusser, L., & Balibar, E. (1970). *Reading capital*. London: New Left Books.

Aristotle (350 BCE). *Metaphysics. Book IV*. Translated by W. D. Ross. Provided by the internet classics archive. http://classics.mit.edu//Aristotle/metaphysics.html. Accessed 24 March 2014.

Bahn, P. G. (2007). *Cave art: A guide to the decorated ice age caves of Europe*. London: Frances Lincoln.

Baldwin, J. M. (1896a). A new factor in evolution. *The American Naturalist, 30,* 441–451, 536–553.

Baldwin, J. M. (1896b). Heredity and instinct. *Science, 3,* 438–441, 558–561.

Baldwin, J. M. (1896c). Physical and social heredity. *American Naturalist, 30,* 422–428.

Baldwin, J. M. (1906). Infant and race psychology (Chapter 1). In J. M. Baldwin (Ed.), *Mental development in the dhild and the race* (3rd ed., pp. 1–33). New York: Macmillan.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1,* 164–180.

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Palo Alto: Stanford University Press.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bohr, N. (1937). Causality and complementarity. *Philosophy and Science, 4,* 289–298.

Bradbury, J. (2003). Human epigenome project-up and running. *PLoS Biology, 1*(3), e82.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge: Harvard University Press.

Bruner, J. (1986). *Actual minds, possible worlds*. New York: Plenum.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: The University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). New York: University of California Press.

Bühler, K. (1934/1982). *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Stuttgart: UTB Gustav Fischer.

Call, J., & Tomasello, M. (Eds.). (2007). *The gestural communication of apes and monkeys*. Manhaw: Lawrence Erlbaum Associates.

Caprara, F. (1996). *The web of life: A new scientific understanding of living systems*. New York: Anchor Books.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M. A., Laureys, S., Tononi, G., & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine, 5,* 198ra105.

Cassirer, E. (1942/2011). *Zur Logik der Kulturwissenschaften. Fünft Studien*. Hamburg: Felix Meiner.

Collingwood, R. G. (1940). *An essay on metaphysics*. Oxford: Clarendon.

Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology, 98,* 448–459.

Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Company.

Deutscher, G. (2010). *Through the language glass: Why the world looks different in other languages*. New York: Henry Holt and Company.

Diriwächter, R., & Valsiner, J. (Eds.). (2008). *Striving for the whole: Creating theoretical syntheses*. Somerset: Transaction.

Dobzhansky, T. (1972). Genetics and the diversity of behavior. *American Psychologist, 27,* 523–530.

Durkheim, E. (1895/1982). *The rules of the sociological method* (*Les règles de la méthode sociologique*, translated by W.D. Halls). New York: The Free Press.
Durkheim, E. (1919). *Les règles de la méthode sociologique* (7th ed.). Paris: Libraire Felix Alcan.
Fahrenberg, J. (1979). Das Komplementaritätsprinzip in der psychophysiologischen Forschung und psychosomatischen Medizin. *Zeitschrift für Klinische Psychologie und Psychotherapie, 27,* 151–167.
Fahrenberg, J. (1992). Komplementarität in der psychophysiologischen Forschung. Grundsätze und Forschungspraxis. In E. P. Fischer, H. S. Herzka, & K. H. Reich (Eds.), *Widersprüchliche Wirklichkeit. Neues Denken in Wissenschaft und Alltag. Komplementarität und Dialogik* (pp. 43–77). München: Piper.
Fahrenberg, J. (2004). *Annahmen über den Menschen. Menschenbilder aus psychologischer, biologischer, religiöser und interkultureller Sicht*. Heidelberg-Kröning: Asanger-Verlag.
Fahrenberg, J. (2008a). Die Wissenschaftskonzeptionen der Psychologie bei Kant und Wundt als Hintergrund heutiger Kontroversen. Struktureller Pluralismus der Psychologie und Komplementaritätsprinzip. Defizite der Philosophischen und Psychologischen Anthropologie und ein Plädoyer für eine interdisziplinäre Anthropologie. *e-Journal Philosophie der Psychologie, 10,* 1–35. http://www.jp.philo.at/texte/FahrenbergJ2.pdf.
Fahrenberg, J. (2008b). Gehirn und Bewusstsein. Neurophilosophische Kontroversen. In S. Gauggel & M. Herrmann (Eds.), *Handbuch der Neuro- und Biopsychologie* (pp. 28–43). Göttingen: Hogrefe.
Fahrenberg, J. (2013). *Zur Kategorienlehre der Psychologie. Komplementaritätsprinzip. Perspektiven und Perspektiven-Wechsel*. Lengerich: Pabst Science.
Faucheux, C. (1976). Cross-cultural research in experimental social psychology. *European Journal of Social Psychology, 6,* 269–322.
Fischer, D. H. (1970). *Historians' fallacies: Toward a logic of historical thought* (pp. 209–213). New York: Harper Torchbooks.
Fraisse, P. (1964). *The psychology of time*. London: Eyre & Spottiswoode.
Freud, S. (1915). The unconscious. *Standard Edition, 14,* 159–214.
Gaver, W. (1996). Affordances for interaction: The social is material for design. *Ecological Psychology, 8,* 111–129.
Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
Geertz, C. (1988). *Works and lives: The anthropologist as author*. Cambridge: Polity.
Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology, 26,* 309–320.
Gernhardt, A., Rübeling, H., & Keller, H. (2013). "This is my family": Differences in children's family drawings across cultures. *Journal of Cross-Cultural Psychology, 44,* 1166–1183.
Gibson, J. J. (1967). New reasons for realism. *Synthese, 17,* 162–172.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gifford, R. (1997). *Environmental psychology. Principles and practice* (2nd ed.). Boston: Allyn and Bacon.
Gillespie, A. (2006). *Becoming other: from social interaction to self-reflection.* IAP - Information Age Publishing.
Gillespie, A., & Zittaun, T. (2010). Studying the moment of thought. In A. Toomela & J. Valsiner (Eds.), *Methodological thinking in psychology: 60 years gone astray?* (pp. 69–88). Charlotte: Information Age.
Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik, 38,* 173–198.
Gumert, M. D., Kluck, M., & Malaivijitnond, S. (2009). The physical characteristics and usage patterns of stone axe and pounding hammers used by long-tailed macaques in the Andaman Sea region of Thailand. *American Journal of Primatology, 71,* 594–608.
Haas, V. (2004). Les cartes cognitives: Un outil pour étudier la ville sous ses dimensions sociohistoriques et affectives. *Bulletin de Psychologie, 474,* 621–633.
Hall, E. T. (1966). *The hidden dimension*. Garden City: Anchor.

Harré, R. (1983). *Personal being: A theory for individual psychology*. Cambridge: Harvard University Press.
Hartmann, N. (1964). *Der Aufbau der realen Welt. Grundriss der allgemeinen Kategorienlehre* (3rd ed.). Berlin: Walter de Gruyter.
Hegel, G. W. F. (1807). *System der Wissenschaft. Erster Teil. Die Phänomenologie des Geistes*. Bamberg: Goebhardt. https://archive.org.
Hirschberger, J. (1980a). *Geschichte der Philosophie. Band I. Altertum und Mittelalter*. Frankfurt a. M.: Zweitausendeins.
Hirschberger, J. (1980b). *Geschichte der Philosophie. Band II. Neuzeit und Gegenwart*. Frankfurt a. M.: Zweitausendeins.
Hoche, H. U. (2008). *Anthropological complementarisms. Linguistic, logical, and phenomenological studies in support of a third way beyond dualism and monism*. Paderborn: Mentis.
Holloway, R. L. Jr. (1969). Culture: A human domain. *Current Anthropology, 10,* 47–64.
Jablonka, E., & Lamb, M. (2005). *Evolution in four dimensions—Genetic, epigenetic, behavioral, and symbolic Variation in the history of life*. Cambridge: MIT Press.
James, W. (1890). *Principles of psychology* (Vol. 1). New York: Holt.
Kant, I. (1781/1998). *Kritik der reinen Vernunft* (Hrsg. J. Timmermann). Hamburg: Felix Meiner Verlag.
Kant, I. (1786/1968). *Metaphysische Anfangsgründe der Naturwissenschaft* (Hrsg. B. Erdmann; P Menzer, & A. Höfler). Preussische Akademie der Wissenschaften, Textausgabe Band IV (pp. 465–565). Berlin: Walter de Gruyter.
Kant, I. (1798/2000). *Anthropologie in pragmatischer Hinsicht* (Hrsg. R. Brandt). Hamburg: Felix Meiner Verlag.
Kellert, S. H. (1993). *In the wake of chaos: Unpredictable order in dynamical systems*. Chicago: University of Chicago Press.
Kelly, G. A. (1955). *The psychology of personal constructs* (Vols. 1 and 2). New York: Norton.
Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace, & World.
Köhler, W. (1969). *The task of Gestalt psychology*. Princeton: Princeton University Press.
Køppe, S. (2012). A moderate eclecticism: Ontological and epistemological issues. *Integrative Psychological and Behavioral Science, 46,* 1–19.
Kuhn, T. S. (1976). The function of dogma in scientific research. In I. R. Averill (Ed.), *Patterns of psychological thought* (pp. 47–66). Washington, DC: Hemisphere Publishing Co.
Lahlou, S. (1998). *Penser-manger*. Paris: Les Presses Universitaires de France.
Lahlou, S. (2001). Functional aspects of social representations. In K. Deaux & G. Philogene (Eds.), *Representations of the social: Bridging theoretical traditions* (pp. 131–146). Oxford: Blackwell.
Lahlou, S. (2008). *L'Installation du Monde: De la représentation à l'activité en situation*. Aix-en-Provence, Université de Provence, Habilitation à Diriger des Recherches en Psychologie, 375.
Lahlou, S. (2011). How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social Science Information, 50,* 607–655.
Laureys, S., Pellas, F., Van Eeckhout, P., Ghorbel, S., Schnakers, C., Perrin, F., Berre, J., Faymonville, M. E., Pantke, K. H., Damas, F., Lamy, M., Moonen, G., & Goldman, S. (2005). The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research, 150,* 495–511.
Le Poidevin, R. (2004). A puzzle concerning time perception, *Synthese,* 142, 109–42.
Le Poidevin, R. (2011). The experience and perception of time. The Stanford Encyclopedia of Philosophy (Fall 2011 Edition), Edward N. Zalta (Ed.). http://plato.stanford.edu/archives/fall2011/entries/time-experience/. Accessed 13 Jan 2013.
Levine, D. (Ed.). (1971). *Simmel: On individuality and social forms*. Chicago: Chicago University Press.
Levine, J. (2003). Experience and representation. In. Q. Smith & A. Jokic (Eds.), *Consciousness: New essays* (pp. 121–136). Oxford: Oxford University Press.
Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.

Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
Li, S.-C. (2003). Biocultural orchestration of developmental plasticity across levels: The interplay of biology and culture in shaping the mind and behavior across the life span. *Psychological Bulletin, 129,* 171–194.
Locke, J. (1689). Essay concerning human understanding. Book I. The Project Gutenberg EBook #10615. http://www.gutenberg.org. Accessed 8 Sept 2013.
Luisi, P. L. (2003). Autopoiesis: A review and a reappraisal. *Die Naturwissenschaften, 90,* 49–59.
Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50,* 370–396.
Mayr, E. (1966). *Animal species and evolution*. Cambridge: Belknap.
Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge: Harvard University Press.
McAdams, D. (2001). The psychology of life stories. *Review of General Psychology, 5,* 100–122.
McGrew, W. C. (1992). *Chimpanzee material culture: Implications for human evolution*. Cambridge: Cambridge University Press.
Mead, G. H. (1934). *Mind, self, and society*. Chicago: The University of Chicago Press.
Mellor, D. H. (1985). *Real time*. Cambridge: Cambridge University Press.
Merleau-Ponty, M. (1967). *The structure of behaviour*. Boston: Beacon.
Millikan, R. (1993). *White queen psychology and other essays for Alice*. Bradford: MIT Press.
Moscovici, S. (1961). *La psychanalyse, son image et son public*. Paris: PUF. (Published in English as Moscovici, S. (2008). *Psychoanalysis, its image and its public*. Cambridge, UK: Polity Press).
Ogden, C. K. (1932). *Bentham's theory of fictions*. New York: Harcourt Brace.
Omi, Y. (2012). Tension between the theoretical thinking and the empirical method: Is it an inevitable fate for psychology? *Integrative Psychological and Behavioral Science, 46,* 118–127.
Pauli, R. (1927). *Einführung in die experimentelle Psychologie*. Leipzig: Quelle und Meyer.
Peirce, C. S. (1902/1958). The simplest mathematics (CP 4.227-323). In *Collected papers of Charles Sanders Peirce*, Vols. 1–6, C. Hartshorne & P. Weiss (Eds.), vols. 7–8, A. W. Burks (Ed.), (1931–1935/1958–1966). Cambridge: Belknap Harvard University Press.
Peirce, C. S. (1931–1935/1958–1966). The collected papers of Charles Sanders Peirce (Vols. 1–6), C. Hartshorne & P. Weiss (Eds.), Vols. 7–8, A. W. Burks (Ed.), (1931–1935/1958–1966). Cambridge: Belknap Harvard University Press.
Pellegrini, A. D. (2011). "In the eye of the beholder": Sex bias in observations and ratings of students' aggression. *Educational Researcher, 40,* 281–286.
Piaget, J. (1928/1995). Genetic logic and sociology. In J. Piaget (Ed.), *Sociological studies*. New York: Routledge.
Piaget, J. (1969). *The child's perception of time*. London: Routledge & Kagan Paul.
Prigogine, I. (1996). *The end of certainty. Time, chaos, and the new laws of nature*. New York: The Free Press.
Ramón y Cajal, S. F. (1897/1999). *Advice for a young investigator* (Translation by Neely Swanson and Larry W. Swanson. Original *Reglas y consejos sobre investigación cientifica: los tónicos de la voluntad*). Cambridge: MIT Press.
Rogers, C. (1959). A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science. Vol. 3: Formulations of the person and the social context*. New York: McGraw Hill.
Rosenzweig, M. R., Leiman, A. L., & Breedlove, S. M. (1999). *Biological psychology: Introduction to behavioral, cognitive, and clinical neuroscience* (2nd ed.). Sunderland: Sinauer Associates.
Rothschuh, K. E. (1963). *Theorie des Organsimus. Bios – Psyche – Pathos* (2. erw. Aufl.). München: Urban & Schwarzenberg.
Rotter, J. B. (1954). *Social learning and clinical psychology*. New York: Prentice-Hall.
Rotter, J. B. (1981). The psychological situation in social learning theory. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective*. Hillsdale: Lawrence Erlbaum.
Royce, J. (1891). *The religious aspect of philosophy: A critique of the bases of conduct and of faith*. Boston: Houghton Mifflin.

Russell, B. (1927). *An outline of philosophy*. London: George Allen & Unwin.
Salvatore, S., Valsiner, J., Gennaro, A., & Simon Travers, J. B. (Eds.). (2010). *Yearbook of idiographic science* (Vol. 2). Rome: Firera & Liuzzo Group.
Sato, T., Wakabayashi, K., Nameda, A., Yasuda, Y., & Watanabe, Y. (2010). Understanding a personality as a whole. Transcending the Anglo-American methods focus and Continental-European holism through a look at dynamic emergence processes. In A. Toomela & J. Valsiner (Eds.), *Methodological thinking in psychology: 60 years gone astray?* (pp. 89–119). Charlotte: Information Age.
Schacter, D. L., & Addis, D. R. (2007). Constructive memory: Ghosts of past and future. *Nature, 445,* 27.
Schrödinger, E. (1944). *What is life? The physical aspect of the living cell*. Cambridge: Cambridge University Press.
Schrödinger, E. (1951). *Science and humanism: Physics in our time*. Cambridge: University Press Cambridge.
Schrödinger, E. (1958). *Mind and matter*. Cambridge: Cambridge University Press.
Schwarz, M. (2014). The living fossil of human judgment. The living fossil of human judgment. *Integrative Psychological and Behavioral Science, 48,* 211–237.
Searle, J. R. (2003). *Rationality in action*. Cambridge: MIT Press.
Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science, 210*(4471), 801–803.
Sherrington, C. S. (1906). *The integrative action of the nervous system*. New Haven: Yale University Press.
Sherrington, C. S. (1940). *Man on his nature*. Cambridge: University Press Cambridge.
Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intra-individual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67,* 674–687.
Shotter, J. (1975). *Images of man in psychological research*. London: Methuen.
Shweder, R. A., & Sullivan, M. A. (1990). The semiotic subject of cultural psychology. In L. A. Pervin (Ed.), *Handbook of personality* (pp. 399–416). New York: Guilford.
Simmel, G. (1908). *Soziologie. Untersuchungen über die Formen der Vergesellschaftung* (1st ed.). Berlin: Duncker & Humblot.
Simmel, G. (1919). Der Begriff und die Tragödie der Kultur. In G. Simmel (Ed.), *Philosophische Kultur* (2nd ed., pp. 223–253). Leipzig: Alfred Kröner. http://socio.ch/sim/phil_kultur/kul_13.htm. Accessed 13 Oct 13.
Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs: Prentice-Hall.
Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Knopf.
Sprung, L., & Sprung, H. (1984). *Grundlagen der Methodologie und Methodik der Psychologie*. Berlin: Deutscher Verlag der Wissenschaften.
St. Augustine. (397/1955). *Confessions and enchiridion* (transl. and Ed. by A. C. Outler). Philadelphia: Westminster.
Stent, G. S. (1969). *The coming of the Golden Age: A view of the end of progress*. Garden City: The Natural History Press. (The American Museum of Natural History).
Stern, W. (1924). *Wertphilosophie* (*Person und Sache. System des kritischen Personalismus. Dritter Band*). Leipzig: Johann Ambrosius Barth.
Teo, T., & Febbraro, A. (2003). Ethnocentrism as a form of intuition in psychology. *Theory and Psychology, 13,* 673–694.
Thomae, H. (1988). *Das Individuum und seine Welt* (2. völlig neu bearb. Aufl.). Göttingen: Hogrefe.
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55,* 189–208.
Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
Tomasello, M. (2009). *Why we cooperate*. Cambridge: MIT Press.
Tomasello, M. (2014). *A natural history of human thinking*. Cambridge: Harvard University Press.

Toomela, A. (2008). Variables in psychology: A critique of quantitative psychology. *Integrative Psychological and Behavioral Science, 42,* 245–265.
Toomela, A. (2011). Travel into a fairy land: A critique of modern qualitative and mixed methods psychologies. *Integrative Psychological and Behavioral Science, 45,* 21–47.
Toomela, A. (2012). Guesses on the future of cultural psychology: Past, present, and past. In J. Valsiner (Ed.), *Oxford handbook of culture and psychology* (pp. 998–1033). New York: Oxford University Press.
Toomela, A., & Valsiner, J. (Eds.) (2010). *Methodological thinking in psychology: 60 years gone astray?* Charlotte: Information Age.
Uher, J. (2008a). Three methodological core issues of comparative personality research. *European Journal of Personality*, *22,* 475–496.
Uher, J. (2008b). Comparative personality research: Methodological approaches. *European Journal of Personality*, *22,* 427–455.
Uher, J. (2011a). Individual behavioral phenotypes: An integrative meta-theoretical framework. Why 'behavioral syndromes' are not analogues of 'personality'. *Developmental Psychobiology, 53,* 521–548.
Uher, J. (2011b). Personality in nonhuman primates: What can we learn from human personality psychology? In A. Weiss, J. King, & L. Murray (Eds.), *Personality and Temperament in Nonhuman Primates* (pp. 41–76). New York: Springer.
Uher, J. (2013). Personality psychology: Lexical approaches and assessment methods reveal only half of the story. A metatheoretical analysis. *Integrative Psychological and Behavioral Science, 47,* 1–55.
Uher, J. (2015a). Conceiving "personality": Psychologists' challenges and basic fundamentals of the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. *Integrative psychological and Behavioral Science, 49*. DOI: 10.1007/s12124-014-9283-1.
Uher, J. (2015b). Developing "personality" taxonomies: Metatheoretical and methodological rationales underlying selection approaches, methods of data generation and reduction principles. *Integrative Psychological and Behavioral Science, 49*. DOI: 10.1007/s12124-014-9280-4.
Uher, J. (2015c). Interpreting "personality" taxonomies: Why previous models cannot capture individual-specific experiencing, behaviour, functioning and development. Major taxonomic tasks still lay ahead. *Integrative Psychological, Behavioral Science, 49*. DOI: 10.1007/s12124-014-9281-3.
Uher, J. (2016). Exploring individuals and the workings of their minds: Methodological principles derived from the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. *Annals of Theoretical Psychology*, *13*.
Uher, J., Addessi, E., & Visalberghi, E. (2013a). Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (*Cebus apella*). *Journal of Research in Personality, 47,* 427–444.
Uher, J., Werner, C. S., & Gosselt, K. (2013b). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality, 47,* 647–667.
Valenti, S. S., & Good, J. M. M. (1991). Social affordances and interaction I: Introduction. *Ecological Psychology, 3,* 77–98.
Valsiner, J., & Han, G. (2008). Where is culture within the dialogical perspectives on the self? *International Journal for Dialogical Science, 3,* 1–8.
Valsiner, J. (1987). *Culture and the development of children's actions: A cultural-historical theory of developmental psychology*. New York: Wiley.
Valsiner, J. (2000). *Culture and human development*. London: Sage.
Valsiner, J. (2012). *A guided science: History of psychology in the mirror of its making*. New Brunswick: Transaction.
Valsiner, J. (2014). Breaking the arrows of causality: The idea of catalysis in its making. In K. R. Cabell & J. Valsiner (Eds.), *The catalysing mind. Annals of theoretical psychology* (Vol. 11, pp. 17–33). New York: Springer.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems, 5,* 187–196.
Visalberghi, E., & Fragaszy, D. (2012). What is challenging about tool use? The capuchin's perspective. In T. R. Zentall & E. A. Wasserman (Eds). *The Oxford handbook of comparative cognition*, Chapter 40. Oxford: Oxford University Press.
von Bertalanffy, L. (1973). *General system theory: Foundations, development, applications*. New York: George Braziller.
von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Springer.
von Uexküll, J. (1980). *Kompositionslehre der Natur. Biologie als undogmatische Naturwissenschaft. Ausgewählte Schriften*. (Edited by Thure von Uexküll). Frankfurt a. M.: Verlag Ullstein GmbH.
Vygotsky, L. S. (1934/1962). *Thought and language*. Cambridge: MIT Press.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
Vygotsky, L., & Luria, A. (1930/1994). Tool and symbol in child development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 99–174). Oxford: Blackwell.
Waddington, C. H. (1942). Canalisation of development and the inheritance of acquired characters. *Nature, 150,* 563–564.
Wagoner, B. (2009). The experimental methodology of constructive microgenesis. In J. Valsiner, P. Molenaar, N. Chaudhary, & M. Lyra (Eds.), *Handbook of dynamic process methodology in the social and developmental sciences* (pp. 99–121). New York: Springer.
Wallace, A. R. (1858). On the tendency of varieties to depart indefinitely from the original type. In Linnean Society of London (1908). *The Darwin-Wallace celebration held on Thursday, 1st July, 1908* (pp. 98–107). London, UK: Burlington House, Piccadilly.
Walach, H. (2013). *Psychologie. Wissenschaftstheorie, philosophische Grundlagen und Geschichte*. Ein Lehrbuch (3., überarb. Auflage). Unter Mitarbeit von N. v. Stillfried. Stuttgart: Kohlhammer.
Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158–177.
Watson, J. D., & Crick, F. (1953). A structure for desoxyribose nucleic acid. *Nature, 171,* 737–738.
Weber, M. (1904). Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik, 19,* 22–87.
Weber, M. (1922/1978). *Economy and society: An outline of interpretive sociology*. Berkley: University of California Press
Weber, M. (1930). *The Protestant ethic and the spirit of capitalism* (trans: Talcott Parson). London: George Allan & Unwin.
Weber, M. (1946). *From Max Weber: Essays in sociology*. (Translated and edited by H. H. Gerth & C. W. Mills). New York: Oxford University Press.
Weber, M. (1949). *The methodology of the social sciences* (Translated and edited by E. A. Shils & H. A. Finch). New York: Free Press.
Westen, D. (1996). A model and a method for uncovering the nomothetic from the idiographic: An alternative to the Five-Factor Model. *Journal of Research in Personality, 30,* 400–413.
Whitehead, A. N. (1929). *Process and reality*. New York: Harper.
Whiten, A. (Ed.). (1991). *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. Oxford: Basil Blackwell.
Wolpert, L. (1992). *The unnatural nature of science*. London: Faber.
Wundt, W. (1863). *Vorlesungen über die Menschen- und Thierseele*. Hamburg: Voss.
Wundt, W. (1894). Über psychische Kausalität und das Prinzip des psycho-physischen Parallelismus. *Philosophische Studien, 10,* 1–124.
Wundt, W. (1896). *Grundriss der Psychologie*. Stuttgart: Körner. Online at https://archive.org/. Accessed 26 Jul 2014.
Wundt, W. (1904). *Principles of physiological psychology*. London, UK: Allen.
Zeleny, M. (1977). Self-organization of living systems: A formal model of autopoiesis. *International Journal of General Systems, 4,* 13–28.

Chapter 14
Agency and Creativity in the Midst of Social Change

Sarah H. Awad and Brady Wagoner

The question of agency has long been a stumbling block for psychology. After a promising beginning in the first decades of the twentieth century with the Würzburg school, Vygotsky, and others, psychology has tended to take the route of denying agency. What we are now given in most psychological studies are analyses that show how one thing correlates with another, how a certain stimulus causes a response, or how a particular input leads to an output. In all this, we find a one-sided focus on isolated "lower" forms of behavior so as to more easily predict some outcome, at least at the level of a population (Molenaar 2004). Agency is the property of individuals acting within a social and cultural world, and calls for a temporal analysis. The analysis of populations, through "interindividual" variations, loses sight of individuals and their becoming, and with it the notion of agency. What is missing is a genuine look at "higher" mental processes, in which a systemic agent—an active person or group—purposefully constructs meaning in order to act in their world. This dynamic and creative process is by definition unpredictable, but is nevertheless constrained by a number of factors that the researcher can identify and study. Unpredictability here is not seen as an obstacle but rather an opportunity to explore the individual and sociocultural factors that facilitate its emergence or those that constrain it.

The present chapter makes use of a sociocultural approach to agency with its attention to "higher" mental functions and the construction of genuinely new cultural forms. In this approach, the focus is on agency through the construction of "signs," which guide ones action into the future (Valsiner 2003, 2007). The classic example of this is discussed in relation to the philosophical problem of Buridan's ass: A donkey having to decide between two equally good barrels of hay dies of hunger. Following Spinoza's theorizing, Lev Vygotsky (1987) argues that the human being will unbalance the options by introducing a sign into the situation—for example, rolling a dice in order to decide between two things. In one way, this may seem to be the opposite of agency, because the way the dice lands makes the decision for

S. H. Awad (✉) · B. Wagoner
Aalborg University, Aalborg, Denmark
e-mail: sarahassanko@gmail.com, awads@hum.aau.dk

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_14

the person. But we should not locate agency in the moment of action so much as the moment previous to it. Agency should be located in the process of sign construction that precedes an action and will at the next moment guide action.

To take another example, Tamara Dembo (1925) did an experiment in which he brought participants into a room, where they were told to wait. They continued to wait for the experimenter who did not return. The question becomes at what point do they get up and leave, and by what means will they decide to do so. Dembo found that participants would look at the clock and say to themselves "when the hand reaches point X, I will get up and leave," which they did. Again, the act of volition here comes before the action is performed in the setting up of a sign in the environment, which latter acts back on them, stimulating them to get up and leave. The construction of a sign thus redefines the situation, allowing the person to imaginatively project themselves beyond the here-and-now, thereby opening up different possibilities for action. In what follows, we will discuss this as a form of "as if" thinking, which is essential for the redefinition of social reality.

Although these examples are of individual action, it is important to stress that sign construction is part of a social and cultural process. As Vygotsky (1987) famously said, all higher mental functions begin as actual relations between people (intermentally) and are internalized only later, so that they function for the person herself (intramentally). Language, for example, is a social product that fundamentally transforms a person's thinking when it is internalized. Although we learn to speak the "same" language, we give it our own accent, idiosyncratic expressions, and uses. Signs are not merely shared but also creatively manipulated, synthesized, transformed, and objectified in a novel form for new uses. We thus need to attend to the cyclical process of experiencing signs objectified by others and how they are interpreted and recreated by individuals to perform an action and construct social reality. Signs are both the means and outcome of "higher" action. In organized groups, these dynamics can lead to a situation of flow and the flourishing of new and creative expressions, which we will demonstrate with a case study of graffiti during the 2011 Egyptian Revolution.

Graffiti in Egypt is looked at as a sign that was constructed in a time of social change as a result of active groups of artists expressing their agency against a definition of social reality that they rejected. The graffiti was a result of the uprising of people who realized the power they had to change society and decided to proclaim their country from a dictatorship. Through their graffiti, artists created signs that express their perspective, promote their revolutionary ideology, and communicate their message within their culture and cross-culturally with people outside of Egypt, transgressing the barrier of language. It is an act of agency in that the graffiti artists creatively constructed new meanings, in opposition to the ruler powers, that can be used as a resource by themselves and others to imagine a new not-yet-existing Egypt.

This case study explores street graffiti as a special kind of sign that was used as a tool by revolutionaries to communicate their message and make an impact on the public. We will look at how revolutionary graffiti emerged as a form of resistance during the 2011 Egyptian Revolution, bringing underground artists to the surface in a collaborative effort. The phenomenon is studied from a creativity perspective

discussing what characteristics support seeing this art as an expression of agency through group creativity and what social factors facilitated it to come about. The focus will be on its emerging form in the first year after protestors took to the streets on 25 January 2011 and on the artwork centered around Tahrir Square. This will define the scope of the study since the Egyptian Revolution has taken different turns afterwards. We argue that revolutionary graffiti offers an understanding of creative agency as a collaborative group process using imagination to move beyond reality and present a peaceful and liberating form of expression that would not have emerged individually. Thus, the case study of Egyptian graffiti is used to draw attention to agency found in group creativity, contrary to Freud's and Le Bon's tendencies to relate groups to violence and chaos.

14.1 Revolutionary Graffiti in Egypt

Human beings have been making marks on surfaces since early in the species history (Donald 1991). In so doing, they began to saturate their environment with external signs that regulated themselves from the outside. Graffiti is a particular species of sign marking, which refers to unsanctioned writing, drawing, or painting in public spaces. The word "graffiti" comes from the Greek word "γράφειν" which means "to write" or "to inscribe." The term was used to refer to drawings and writings that were scribbled onto ancient walls and temples such as those found from the Pharaonic and Roman remains. In modern time, it is used to refer to street paintings that are usually implemented by anonymous youth who choose graffiti as an alternative form of dialogue to represent themselves to the public. The culture of graffiti has, for a long time, transformed city spaces, such as New York City and Shoreditch in London, and has established a new form of communication using words, symbols, and graphics. Through its history there has been continuous controversy about graffiti's legality and whether it should be supported or penalized. As the graffiti artist Banksy ironically said, "if graffiti changed anything it would be illegal." However, when looking at contemporary international art movements, the art is on the walls (Wacławek 2011).

Graffiti in Egypt could arguably be dated back to Ancient Egypt where Pharaohs documented different aspects of their lives through engraved paintings on walls. In modern day Egypt, artistic expression was mostly seen in cultural centers, art schools, and within underground artists' groups. With the 2011 uprisings came an explosion of artistic activity, including folksongs, drama, and a new form of "mesmerizing" graffiti, as Abaza (2013) describes it. This graffiti started as a revolutionary tool to communicate and mobilize people around the main goals of the revolution. As one artist explains "painting graffiti was one main way of both defending and occupying the street. It was a way of conquering the space in a situation of war" (cited in Abaza 2013). Around this idea formed a large group of artists and amateurs connected by the space in Tahrir Square, where they met under a sign saying "Revolution Artists Union."

Fig. 14.1 Regime-confronting art. (Photo credit: Ranya Habib, Tahrir Square, October 2012)

Their graffiti paintings communicated a variety of themes. The early graffiti incorporated the main slogans of the revolution such as "power to the people" and "bread, freedom, and social justice." Those paintings helped mobilize and unite people around the same message as well as being a tool for expressing the people's demands. Other works included caricatures of the ruling generals, mocking and directly insulting the Mubarak regime, and later the Supreme Council of the Armed Forces (SCAF). Thus, graffiti functioned as a sign to claim public space, articulate values, remember events, create solidarity, empower and mobilize people, and critique powerful institutions and individuals.

The photo of the graffiti in Fig. 14.1 was taken in October 2011 at Tahrir Square. It illustrates a painting on the left side of former president Hosni Mubarak and Mohamed Hussein Tantawi, the chairman of the SCAF, who was in charge after the ouster of Mubarak. Both their faces are painted in a way to express how they are two sides of the same coin. Behind their face the photo of Mohamed Badie was added after Mohamed Morsi became the president in June 2012, Badie was the Supreme Guide of the Muslim Brotherhood and was known to be the one behind all Morsi's decisions while in power. The text underneath reads, "the commander

never dies" rhyming with a local proverb that expresses how offspring are similar to their parents so in a way their parents never die. Similarly, the graffiti expresses how Mubarak's regime and corruption still lives on with the succession of Tantawi and Morsi. The right side of the graffiti illustrates an artist dressed in the Egyptian flag colors confronted by authority. Underneath it says:

> A regime fearing a painting brush and a pen
> An unjust system that attacks the victim
> If you were righteous, you wouldn't have feared what I draw
> All you do is fight walls, show off your power on paintings
> But inside you, you are a coward
> You will never rebuild what has been destroyed

The theme in this graffiti as well as many similar ones challenged the barrier of fear towards the regime. The graffiti reflected the power as fearful and protected by the security forces against the simple tools of an artist expressing the weakness of the regime in the face of the creative agency of an individual. The image associates the rulers with violence and repression, and the revolutionaries as peaceful and patriotic in their fight against them. It expresses a storyline constructed in "as if" mode that calls for revolutionary actors to step into the role of artists contra to those in power. Moreover, the image of Tantawi and Mubarak continued to be reproduced, as new figures became leaders and were added in the place of Badie behind the other two. This is an effective strategy for building equivalences between those who have been in power, and thereby defining social reality in such a way that it becomes a duty to continue to fight against authority.

Artists also used graffiti to remember and document major events happening in the square as well as to memorialize and give face to the protestors who died in the revolution. The graffiti in Fig. 14.2 is a painting of Mina Daniel, who died in a protest in front of the state TV. In all these designs we see how the artists used graffiti to transform city space into a setting saturated with signs of their own creation, in contrast to other parts of the city dominated by signs promoting those in power (e.g. portraits of politicians portrayed as patriotic heroes). One cannot avoid hearing their voice over that of those in power when walking by, though not everyone accepted their message, as we will see below. Those different themes have turned the center of the capital, as Aboul Ezz (2012) describes it, into an open-air art gallery of the revolution that reconfigures and subverts state symbols to make their point and mobilizes people under its banner.

Later, as the authorities started building walls around Tahrir Square to protect vital buildings and to limit people's access to the square, artists started to use those same walls to draw scenery, breaking the boundaries and reclaiming the space with imaginary landscapes (Aboul Ezz 2012). Similar graffiti has been created along the walls separating Israel and the West Bank, the USA and Mexico, and other segregating walls. As Abaza (2013) suggests, the artists in this "No Walls" campaign were probably inspired by the work of the British graffiti artist Banksy, who used "trompe l'oeil" to penetrate with imagination segregating walls. The "No Walls" campaign was interesting in how it culturally borrowed a symbol and developed it locally to redefine barriers and confront the government's power by imagination.

Fig. 14.2 Remembering those who died during the revolution: "You're in our hearts, Mina." (Photo credit: Ranya Habib, Tahrir Square, October 2012)

The obstacles built by the government were replaced by signs that visually penetrated them, calling for the protests to do the same in their action (Fig. 14.3).

The graffiti work during this period emerged as part of the social movement of the Egyptian Revolution. It was an effort to record and celebrate its history, as well as to fill the void that the government has continuously failed to fill—for example, paying tribute to the dead, holding the perpetrators accountable, and restoring a sense of normalcy to the changing realities (Morayef 2012). Inspired by the same euphoric spirit of the revolution, the artwork was an attempt at breaking the fear barrier and demanding power to the people. In short, the city became saturated with signs promoting solidarity and empowerment among Egyptians, together with a portrayal of a social reality (e.g. key events, actors, and martyrs of the revolution) that should be fought against.

14.2 Was the Revolutionary Graffiti an Innovative Act of Agency?

The present paper argues that the process of graffiti painting during the Egyptian Revolution was a creative innovation requiring the active agency of artists. In line with Boden's (1996) definition of creativity, graffiti as a form of revolutionary

Fig. 14.3 No walls graffiti. (Photo credit: Jonathan Rashad)

expression was new to the culture as will be explained below; it was an unexpected act that emerged in Tahrir Square as a proclaimed space, and it offered a valuable new space and a voice for resilience and resistance. As one artist put it, "we have launched a non-violent campaign, simply by opening up the walls through drawing" (cited in Aboul Ezz 2012). To illustrate how this process could be considered an innovation, an analogy will be used between this phenomenon and musical performance presented in Sawyer's study (2006) of group creativity. Then we will discuss how this innovation aimed at a social symbolic repair utilizing imagination.

Sawyer's (2006) studies with unstructured improvisational musical groups have led him to identifying three characteristics of group creativity: improvisation, collaboration, and emergence. Similar to jazz improvisation, the graffiti paintings emerged at the moment of encounter as a result of activist and artists uniting in one space and proclaiming it as their space. The "performers" in the graffiti work would start a painting with a certain message and one artist after the other would build upon the original message creating threads of paintings along Tahrir Square's walls. Also, similar to how a final jazz-improvised piece is a group outcome based on the interactional dynamics, the graffiti was a result of group effort that cannot be attributed to one person. Even though some paintings had specific signatures to them, the overall outcome was a result of dynamic collaboration that sometimes involved no direct communication between artists. The last characteristic is that of emergence, where the "whole is greater than the sum of the parts" (Sawyer 2006). In musical performance, the final art piece is a creative outcome that is incomparable to each musician's input added together. This characteristic is the most important

in our discussion of the revolution graffiti. The emergent phenomenon resonates a much bigger idea of using space to build resilience and break barriers, an outcome as a whole much greater than each specific art piece on every wall.

The fact that the graffiti emerged from the social movement of the revolution gave it a certain group spirit that facilitated what Sawyer (2006) refers to as the "group flow." This group spirit bloomed when the unexpected number of people kept increasing in the square united by the same demands and sharing the sense of ownership of the square. At this instance, artists may have realized their talents as their tool for activism. They also coordinated their efforts in such a way as to respect each other's work and build on it in their own.

Another way to look at the agency exercised in the revolution graffiti is in how artists utilized imagination. The messages behind the revolution graffiti were an expression of the frustration with the government, a message that was unacceptable in a society in which only mentions of praise to the ruling family were allowed in the public sphere. Innovations that are considered to be unrealistic or breaking the boundaries within certain societies depend on a particular form of symbolic repair to exist for people to reconfigure the borders of specific semiotic set and use imagination to engage in "as if" thinking or action to create new pathways for action (Zittoun et al. 2013, p. 293). Graffiti was a tool to reconfigure the boundaries of what could and could not be said in a country ruled by dictatorship, reclaiming public space for the revolutionary cause. Also, it was a space for imagining what is beyond reality and expressing the artists' wishes for the future of Egypt. This was done by objectifying values of religious tolerance, critical citizenship, and resistance to brutality through image and text on the wall.

Some would argue that the revolutionary graffiti was not innovative because graffiti existed long before the revolution and thus its use as a revolutionary tool does not qualify it as a new innovation. Others would argue against the uprising as a whole and would see the graffiti as a form of destruction and vandalism. Different reflections upon the graffiti were in themselves a projection of the agency of different groups of people in Egypt and outside Egypt. Perspectives in many instances were influenced by the political stance of the perceivers and how they chose to express their position towards the uprising. However, looking beyond graffiti as a revolutionary tool to their aesthetic, the paintings did introduce a new artistic composition that was a fusion of popular Islamic, Coptic, pharaonic, and universal artistic traditions (Abaza 2013). Frederic Bartlett (1923) argued that this kind of welding together of influences coming from different directions in a move towards a particular future goal is what defined genuinely constructive and creative thought (see also Wagoner 2013a, 2013b). The fusion of traditions found in revolutionary graffiti in Egypt succeeded in capturing local and international attention through a visual message transcending the barrier of language. As illustrated in Fig. 14.1, even without understanding the Arabic text, the painting presents the tension between the government and the artists as well as the different yet similar faces of corruption. Not only can these concepts be understood internationally by people

in different cultural contexts but it can also be reflected upon and related to by individuals living in similar struggles in other countries. Each interpreter reflects upon the graffiti from his or her own position, exercising his or her own agency in making meaning of it.

14.3 Institutionalization and Resistance

As with many social innovations, especially revolutionary forms of expression, Egypt's revolutionary graffiti was not accepted by everyone. This also demonstrates that signs are not reacted to uniformly but can become stimuli for very different actions; the meaning is not in the sign but constructed by the agent viewing it. Resistance to the revolutionary graffiti came mostly from the government and by citizens in support of the government. The authorities' countermovement was to continuously erase the graffiti from the walls of the streets around Tahrir Square, and prevent artists from drawing new ones. The artists' response was to utilize social media, launching a campaign on Facebook to call for the repainting of the walls and reclaiming the space (Abaza 2013). As one artist expresses the resistance, "if they play with walls, we will play with the mind. They put up a wall but we do not see it…you express yourself and at most in a few days or a week it will be removed, then you paint again" (cited in Aboul Ezz 2012). In other cases, government-backed groups altered the graffiti so that it fit their own definition of social reality. In one case, a graffiti illustrating the "Mapero Massacre," where military tanks ran over protestors in Alexandra, was changed so that protestors were simply waving Egyptian flags next to a tank (thus associating the military with popular patriotism). Artists responded by painting a SCAF monster devouring protestors.

Resistance to social change involves members of society, either passively or actively rejecting this change (Van der Zanden 1959). The government resistance was an active one aimed at the immediate removal of the graffiti or in some cases alternation. On the other hand, some members of the society resisted it passively by rejecting it as a form of art or expression and calling it vandalism aiming at disrespecting the authorities. Van der Zanden (1959) presents two possible explanations for resistance in his study of social movements: one is the rejection of change as it creates friction and doubts, and the other comes from vested interests and genuine concerns about the future. In the present case study, the first reason could be attributed to the government's resistance that saw the graffiti as a blunt assault on its authority and the second reason could be attributed to people's fear that this rebellious form of expression would cause more instability.

While the government made every effort to not only un-institutionalize the graffiti but also to erase it from the streets and people's memories, many protestors and their supporters recognized the graffiti as a liberating innovation. Artists made efforts to institutionalize the graffiti through forming the Revolution Artist

Association, through online forums (e.g. on the Facebook group "Revolution Graffiti," "Wall Talk," and "Graffiti the Streets of Egypt") and through the documentation of graffiti in books such as *Revolution Graffiti: Street Art of the New Egypt* (Gröndah 2013) and *Walls of Freedom: Street Art of the Egyptian Revolution* (Hamdy et al. 2012). In this, we see an arms race between government and revolutionaries searching for new ways to counter the innovative actions of the other. Ironically, it is the resistance encountered on the way to achieving one's aims that creates conditions for agents to become truly creative in their search for innovative ways around the problem, as can be seen in the "no walls" campaign and forums to retain graffiti in the face of government destruction. New tools get appropriated and adapted by revolutionary agents to solve problems at hand.

14.4 Sociocultural Factors

Social factors and group interaction surrounding any innovation play a crucial role in its composition (Hennessey 2003). The revolution graffiti emerged in a time of major social change in the Egyptian society. It was in a way facilitated by a dynamic social movement that fostered creativity and agency through revolt (Abaza 2013). Through the revolution, artists were able to stand outside of the reality of their situation through the use of their imagination, projecting new possibilities into the future which challenge those in power. Humans manage to escape being trapped in perceptual fields by creating meanings through the use of signs. Those meanings enable them to stand outside a phenomenon, which is a crucial step for agency and creativity to occur (Glaveanu et al. 2014). This also occurs in the form of novels (e.g. of the utopian genre), music and drama, which objectify future possibilities for a group.

The dynamics of Tahrir Square as a space for this innovation can be best explained through Zittoun et al. (2013, pp. 339–340) description of how streets can be "new arenas of play" for adults. There are many constraints to adults' use of their imagination to express their wishes and fears and for this expression to occur there must be a space for play. Through the Egyptian Revolution, the streets were reshaped by protestors allowing the engagement in imagination and play, transcending the boundaries of physical space so as to open up boundless spaces for possibilities. In other words, revolutionaries could objectify their own vision for Egypt in Tahrir.

Public space was redefined as space of performance, contestation, and debate (Mehrex 2012). This confirms the power of the surrounding space in fostering creativity and innovation. Artists used signs to proclaim space in a situation of conflict. In the graffiti in Fig. 14.4, artists have renamed Mohamed Mahmoud Street in Tahrir Square to "Freedom Eyes" Street in reference to an incident in which security forces have targeted the eyes of protestors with rubber bullets in November 2011.

Fig. 14.4 Renaming the streets surrounding Tahrir Square. (Photocredit: Ranya Habib, Tahrir Square, October 2012)

Underneath the painted street sign it said, "interior security forces not allowed." The space was used and transformed by signs to dictate new rules and redefine the social reality according to the artists' and activists' vision.

14.5 Groups as a Positive Force

As discussed before, revolutionary graffiti involved a collaborative group process utilizing different dynamics. An artist, who identifies himself as part of the "Defying Authorities Through Chalk, Paint, Spray and Stencils" movement, explains the process as follows: "the messages expressed in the graffiti were always in parallel to the events, something happens, groups of activists come together, and the artists step in" (as cited in Aboul Ezz 2012). This brings us to how the different communication channels played a crucial role in the group dynamics. Group communication ranged from face-to-face communication between activists and artists, to online communication using social media, to visual communication between different artists through the artwork itself. As Abaza (2013) describes, the square was filled with

numerous artists who were communicating together the next round of politically subversive graffiti art in response to the erasing of previous ones. They were united by a common conscious struggle to preserve the revolution through their paintings.

This form of positive group force is similar to what Heberle (1949) identified as the characteristics of groups influencing social movements. He characterized them as sharing a consciousness of group identity and solidarity along with an awareness of common sentiments and goals, and an existence of a "we feeling" that helps positive change happen. As noted above, this feeling permeated the geographical space of Tahrir and was fortified in the graffiti itself, renaming of places, and the development of group practices there.

The characterization of the social group developed here contrasts with the negative connotations of groups in the majority of social psychological studies. This conception goes back to the works of Le Bon and Freud, and is characteristic of many contemporary theories, such as de-individuation theorists. These thinkers associate groups with aggression, immorality, and loss of freedom. The individual here is considered a free and rational agent until he or she becomes part of a crowd at which point primitive instincts come to the fore (Greenwood 2004). Le Bon (2002) suggested that there is a mental unity in crowds that influences members' consciousness and makes them feel omnipotent and capable of achieving what is impossible for isolated individuals. As a result, we lose all feelings of individual responsibility in the crowd and the "mass exposes the unconscious of us all, the unconscious foundation that is the same, for everyone is exposed" (Freud 2004). Freud agreed with Le Bon in that individuals in crowds reveal their basic instinctual barbaric nature and their unconscious wishes come to the fore, stripping them of all civilized behavior (Billig 1976). This view presents individuals as free and rational outside the group, but once part of a mass they lose their individual consciousness and agency. This view only attributes acts of chaos and aggression to groups, viewing them as incapable of collaborative and innovative use of signs for a meaningful purpose.

Both views are valid interpretations of some of the dynamics in play when groups of people come together. Freud's and Le Bon's direction of thought could be applied to explain the mass chaos and sexual harassment that were also part of the revolution. However, the present case study proposes putting more attention on examples of positive group dynamics that supports the agency of members within the group, seeing them as catalysts for creative outcomes that might not have been achieved individually. The revolutionary graffiti in Egypt was able to grab the attention locally and internationally by its harmonious dissemination in Tahrir Square and spreading to other areas inside and outside Cairo. Its impact was achieved through a group of people doing the same act at the same period of time and transforming a space of conflict into a colorful landscape of imagination. If the phenomenon was limited to few individual artists, its impact would have been different and it might have failed to achieve the same significance it still does after 3 years from the beginning of the revolution.

14.6 Conclusion: Lessons Learned About Agency and Creativity

In this concluding section, we would like to reflect upon different insights on agency and creativity using this case study. One main lesson learned is that creativity can sometimes be best captured as a process of exercising agency, rather than a final individual outcome. Limiting creativity to individualized outcomes leaves the novelty of many great group processes unrecognized. As Wagoner highlights, it is important to look at creativity as a complex ongoing process, oriented to an open future (Gillespie et al. 2014). For the graffiti discussed in the case study, it is not the final creative product of paintings on walls that matters as an innovation, so much as the creative group process of expressing their agency in responding to new events and redefining social reality.

Also of importance is recognizing group forms of creativity. We often try to attribute group creativity to one person because of what Resnick (1996) refers to as our "centralized mindset." We automatically assume there is a single agentic self. However, the essence of the graffiti—as well as the Egyptian Revolution—was in the fact that it had no one leader. The group dynamic presented in the square formed a euphoric flow of creativity that would not have emerged individually. Therefore, we proposed putting more focus on positive aspects of group agency, rather than simply associating group activity with destruction and disorder. This former emerges through the conscious and dynamic agency of members of a group, creating spaces of imagination in resistance to those who hold positions of power.

Another insight is that an innovation does not need to be institutionalized or accepted by authorities to be considered as an innovation. It is hard to have consensus about any one creative outcome and political powers may put down an innovation just because it does not match their agenda. Considering the current authorities in Egypt, the graffiti might never be accepted, officially documented, or given credit, but still that does not undermine it as an innovation. The graffiti becomes recognized within the particular social field of its producers for its aesthetics and ability to make a public impact, and thus is to a certain extent autonomous from the official accreditation bestowed through state institutionalization. In a way, graffiti valued by the social field of its producers is the inverse of its value as seen from the government's perspective.

Finally, as shown in this study, innovation comes from the unique ability of imagination to transcend the current social reality. It comes from the creator's exemplarily talent to look beyond current struggles and limitations and reflect upon oneself as an agentic factor in the composition of the social reality. This can be seen in the representation of revolutionaries' act of rebellion as a dance of freedom (Fig. 14.5) and in the transformation of bullets into birds (Fig. 14.6). The essence of the creative process involved in the graffiti work was in the artists' ability to see beauty and grace in the resistance movement and to become agents in changing their surrounding social representations, opening up new horizons for the future.

Fig. 14.5 Dance. (Source: Revolution Graffiti 2012)

Fig. 14.6 Peace. (Source: Revolution Graffiti 2012)

References

Abaza, M. (2013). Walls, segregating downtown Cairo and the Mohammed Mahmud street graffiti. *Theory, Culture & Society, 30*(1), 122–139.
Aboul Ezz, M. (2012). "Egypt revolution graffiti brings down barricades." *Daily news Egypt*.
Bartlett, F.C. (1923). *Psychology and primitive culture*. Cambridge: Cambridge University Press.
Billig, M. (1976). *Social psychology and intergroup relations*. London: Academic Press.
Boden, M. A. (Ed.). (1996). *Dimensions of creativity*. Cambridge: MIT Press.
Dembo, T. (1925). Vortrag von Fräulein Dembo, über ihre Versuche, mit anschliessenden Fragen. Im Psychologischen. Institut der Universität, Montag d. 7. Dez. 25. Minutes of a meeting on 7 December 1925.
Donald, M. (1991). *Origins of the modern mind*. Cambridge: Harvard University Press.
Freud, S. (2004). *Mass psychology and other writings*. London: Penguin.
Glaveanu, V., Gillespie, A., & Valsiner, J. (2014). *Rethinking creativity: Contributions from social and cultural psychology*. London: Routledge.
Greenwood, J. (2004). *Disappearance of the social in American Social Psychology*. Cambridge: Cambridge University Press.
Gröndah, M. (2013). *Revolution graffiti: Street art of the new Egypt*. Farnborough: Thames & Hudson Ltd.
Hamdy, B., Karl, D. S., & Eltahawy, M. (2012). *Walls of freedom: Street art of the Egyptian Revolution*. From Here to Fame Publishing. Berlin, Germany.

Heberle, R. (1949). Observations on the sociology of social movements. *American Sociological Review, 14*(3), 346–357.
Hennessey, B. A. (2003). Is the social psychology of creativity really social? Moving beyond a focus on the individual. In P. Paulus & B. Nijstad (Eds.), *Group Creativity: Innovation through Collaboration* (pp. 181–201). New York: Oxford University Press.
Le Bon, G. (2002). *The crowd: A study of the popular mind*. London: Dover.
Mehrez, S. (2012). Watching Tahrir. Global Dialogue Newsletter, International Sociological Association, February, 2(3).
Molenaar, P. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, This time forever. *Measurement, 2*(4), 201–218.
Morayef, S. (2012). The seven wonders of the revolution. *Jadaliyya*. http://www.jadaliyya.com/pages/index/4776/the-seven-wonders-of-the-revolution. Accessed 4 March 2014.
Resnick, M. (1996). Beyond the centralized mindset. *The Journal of the Learning Sciences, 5*(1), 1–22.
Revolution Graffiti—Memory of a Nation (ca. 2012). In Facebook (group page). https://www.facebook.com/groups/GRAFTY. Accessed 10 March 2014.
Sawyer, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music, 34*(2), 148–165.
Valsiner, J. (2003). Beyond social representations: A theory of enablement. *Papers on Social Representations, 12*(7), 1–7.16. Retrieved March 2005, from http://www. psr.jku.at.
Valsiner, J. (2007). *Culture in minds and societies: Foundations of cultural psychology*. New Delhi: Sage.
Van der Zanden, J. W. (1959). Resistance and social movements. *Social Forces, 37,* 312–315.
Vygotsky, L. (1987). *The collected works of L.S. Vygotsky. Vol. 4: The history of the development of higher mental functions*. New York: Plenum Press.
Wacławek, A. (2011). *Graffiti and street art*. Thames & Hudson New York: Thames & Hudson. Print.
Wagoner, B. (2013a). Culture and mind in reconstruction: Bartlett's analogy between individual and group processes. In: A. Marvakis, J. Motzkau, D. Painter, R. Ruto-Korir, G. Sullivan, S. Triliva, & M. Wieser, (Eds.), *Doing psychology under new conditions* (pp. 273–278). Concord: Captus Press.
Wagoner, B. (2013b). A systemic approach to cultural diffusion and reconstruction. In: K. R. Cabell & J. Valsiner (Eds.), *The catalyzing mind: beyond models of causality* (pp. 125–147). New York: Springer.
Zittoun, T., Valsiner, J., Salgado, J., Gonçalves, M. M., Vedeler, D., & Ferring, D. (2013). *Human development in the life course: Melodies of living*. Cambridge: Cambridge University Press.

Chapter 15
From Individual Agency to Co-agency

Vlad Petre Glăveanu

Orwell's 1984 novel vividly evokes the limits of human agency in the face of oppressive, totalitarian social systems. Under the permanent surveillance of Big Brother, control is exercised over the body and actions and, most of all, the mind of Oceania citizens. Not only does the party keep a watchful eye over incipient acts of resistance—such as writing a diary—but also its ambition is much greater: to make agency and resistance impossible by eliminating their very means: Newspeak is a simplified language designed to prevent the mere formulation of agentic, revolutionary thought. The novel follows closely the story of Winston Smith whose struggle against Big Brother is prototypical for the human effort to maintain personal agency and, with it, humanity and dignity. This 'battle' over personal agency is played out in the last, terrifying scenes of torture in which O'Brien attempts to gain full control over the mind of Smith by destroying within it the last vestiges of individuality, rationality, and affection. It is thus the individual and, especially, the *individual mind* that come across as the centre of agency while external forces continuously try to colonise this 'centre' and inscribe within it uniformity and complete dependency. The individual and society (granted, in this case, a totalitarian society taken to its extreme) are at odds with each other and the individual mind appears as both the undeniable source and last bastion of human agency.

This iconic image of what it means to be an agent and to lose this status is by no means a concern of the past. On the contrary, Orwell's imaginary world serves as a grim reminder of the dangers of delegating too many responsibilities and sacrificing individual rights to the state and it resonates, for example, with current debates over surveillance in places like Britain or the legitimacy of totalitarian regimes around the world. At the core of such concerns is the relationship between individuals and the collective, a relationship that raises important queries related to freedom, personal initiative, and our capacity to shape and not only be shaped by our social context. We should note here that the question of agency itself has a long history and the way we answer this question is consequential for our understanding of ourselves

V. P. Glăveanu (✉)
Aalborg Universitet, Aalborg, Denmark
e-mail: vlad@hum.aau.dk

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_15

as human beings, of society, and also of the natural world. Are we truly the agents of our individual and collective destiny or simple marionettes in the hands of instincts, of the unconscious, of social institutions and political powers outside our control? This largely existential question is by no means easy to answer and it is not the purpose of this chapter to engage with it in this form. My focus will be directed rather towards unpacking the different meanings of agency in psychology and connected disciplines and proposing a basic framework for what I call 'co-agency'. In essence, this type of conceptualisation challenges purely individual and mentalist readings of agency that do not fully account for the fact that we are not only constrained by our relations with the material and social world but also *defined as agents* by these very relations.

My own interest in the topic of agency has been sparked by the extensive engagement I developed in the past years with the literature on creativity. At first sight, one might in fact wonder if creativity and agency are not interchangeable terms and, indeed, there are today more and more discussions of 'creative agency' (e.g. Wilf 2011). In a nutshell, my view in this regard is that creative acts bear the mark of agency but we are agents also when we do not act 'creatively', meaning when we do not generate outcomes considered by self and/or others as new or significant (Lubart 2003; Runco 2007). In other words, when operating within a mainstream view of the creative process, agency is not reserved for creative people or those moments in which we are visibly creating something. It is only when adopting a broader understanding of creativity that distributes this phenomenon at the level of everyday interactions between people and objects (Glăveanu 2010) that we are able to bring together agency and creativity under a unitary framework of human action that highlights its emergent, transformational quality. Becoming more and more familiarised with theory and research related to agency, I could not help but notice some deep similarities between how we tend to define agents and (or rather as) creative individuals. A sense of agency originating from the person and being expressive of the self, particularly when the self is contrasted or even opposed to others, to society and culture, is strikingly similar to debates over the individual or social nature of creative action.

In the end, as I argue here, this is a false debate since person and society are not two completely separate entities. People exist within society and society exists within the person. The agency and creativity of the person have their origin precisely in this *relational space* defined by individuals and the world they inhabit. Neither of them are 'entities' or 'qualities' possessed by the person in isolation, but markers of the *encounter* between people, objects, and social institutions. Where do such claims leave though Winston Smith and his struggle against an oppressive social system? Is saying that society populates the mind not a way of enforcing O'Brien's argument that Big Brother needs to replace the individual self? What does it mean to discuss Smith's co-agency in relation to that of his antagonist? As this chapter will demonstrate, (re)writing agency is an exercise that forces us to rethink our definitions of person and society and, more than this, to overcome dichotomic views and the temptation to create hierarchies that subjugate one to the other. It is an exercise in systemic thinking that does not lose sight of the individual,

on the contrary, gains a new appreciation of its role and capacity. At the same time, the agency enhancing and also the agency annihilating forces of various societal arrangements need to be carefully reflected upon and, as such, an ethical dimension inscribed into any attempt to theorise agency. In this ambitious quest, we cannot avoid engaging, from the start, with definitional issues and disciplinary differences. Positioning, as follows, the psychological perspective within this varied landscape will help us appreciate much better its strengths and limitations and consider new theoretical horizons.

15.1 The Meaning(s) of Agency

Unsurprisingly for such a complex concept, the meaning of agency is neither singular nor consensual. Different disciplines operate with various definitions and approaches to agency and these understandings communicate as well with lay conceptions of this notion. Moreover, the problem of agency has not been formulated in the same terms across historical times or geographical places. A millenary tradition of reflecting on whether people possess the freedom to shape their own destiny within philosophy and theology led scholars to a reflection on the relation between free will and determinism, the individual and collective aspects of agency, and the need to enhance or control this quality. Today, we live in a day and age in which the question of agency is becoming more and more pressing, and leads people around the globe to question the political class, the premises of dominant economic systems, or our relationship to the environment. New issues such as personal and societal responsibility, as well as a belief in the power of agency to transform human life and society, are evermore present. And yet, the question still remains of what exactly do we mean by agency when we try to understand, discover, protect, or enhance it?

One way to approach this generous question is to consider the etymological origins of this term. The dictionary[1] points us to the Latin concept of *agentem* (*agentia* in Medieval Latin) and its connections to being active, effective, and powerful. Interestingly, more recent references to this term from the nineteenth century designate an establishment where business is done for another. This meaning is preserved in current usage whenever we refer to organisations, companies, or governmental bureaus as agencies. While this institutional reading of the concept does not concern me here, it is important to note that the general idea of acting on behalf of others is fundamental for agency theory as developed in economics and sociology (see Shapiro 2005), where the focus often falls on principle–agent relations. Anthropological accounts, on the other hand, tend to emphasise action and, in particular, discuss how we either act or are acted upon by others and objects (Gell 1998). In philosophy, the idea of agency is closely connected to debates of long duration between partisans of free will and those who see it as constrained, limited, or even illusory. Between existentialist ideas claiming that we are condemned to be free (and responsible for

[1] Online etymology dictionary, available from http://etymonline.com/.

this freedom) and structuralist views there is a wide spectrum of positions that continue to shape debates in the humanities and the social sciences. The controversy over the nature and power of agency is pervasive and has permeated many scholarly domains including organisational studies (Reed 1988), law and policy (Cheliotis 2006), international relations (Hollis and Smith 1994), human geography (Pile 1993), education (Gough 1999), and computing (Friedman and Kahn 1992). Of particular interest remain both traditional accounts of agency in theology (see Browning Helsel; Chap. 10, this volume) and more recent discussions in neuroscience (Lavazza and De Caro 2010; Klemm; Chap. 4, this volume and Zhu; Chap. 5, this volume). Notable for the latter are the contentious findings of Libet (2002) that question the idea of free will on neurological grounds.

While one would expect a thriving area of agency studies in psychology, considering the connection referred to before between agentic qualities and the individual, particularly the mind of the individual, it is surprising to see relatively few discussions of agency among psychologists. This state of affairs can be explained in different ways. Under the influence of behaviourism and, later on, cognitivism, most psychologists tend nowadays to consider agency a rather mysterious, philosophical concept, one that is difficult to operationalise in empirical research. Regrettably, the same kind of theoretical and methodological narrowness led many psychologists in the last decades to become less concerned precisely with the kind of notions that were fundamental for the discipline in its early days. Consciousness is here an example among many. While general psychology textbooks rarely mention agency, this concept remains important for social, cultural, and developmental psychologists. One of the most visible examples of psychological literature on the topic of agency can be found in the writings of Alfred Bandura for whom:

> To be an agent is to influence intentionally one's functioning and life circumstances. In this view, personal influence is part of the causal structure. People are self-organizing, proactive, self-regulating, and self-reflecting. They are not simply onlookers of their behavior. They are contributors to their life circumstances, not just products of them. (Bandura 2006, p. 164)

What is distinctive for Bandura's account, as well as much social psychological thinking about agency, is the emphasis on the reciprocal relation between person and context. Taking stock of past conceptions, Bandura (1989) distinguishes between three distinct conceptualisations defined as autonomous agency, mechanical agency, and emergent interactive agency. Both autonomous and mechanical views of agency are limited by their emphasis on the complete independence of agents and on stimulus–response causal links, respectively. On the contrary, his social cognitive theory stresses the emergent and interactive nature of agency where 'persons are neither autonomous agents nor simply mechanical conveyers of animating environmental influences' (p. 1175). While taking into account both individual and environmental influences, the core of Bandura's theory remains his well-known notion of *self-efficacy*. In his view, it is important to understand not only how external influences impact the thoughts and actions of the person but also how the person's own reflections shape this dynamic. These self-generated influences are a crucial contributing factor and, in fact, offer the central characteristic of human agency.

Bandura's discussion of agency is therefore repeatedly framed in terms of feedback, the representation of outcomes and goals, and forethought. Taking the example of the latter, we are invited to think about how purposive behaviour is regulated by our capacity to anticipate the outcomes of our action. This capacity motivates (or discourages) the person to demonstrate agency by pursuing a certain course of action and not others. Self-directiveness and self-control are key to a cognitive perspective on agency and raise questions of whether and how we are actually departing from the traditional focus on the individual and on internal processes. The simple stimulus–response schema of behaviourist psychology is updated by adding stimuli generated by the person himself or herself but this does not account for the complexity of the world outside the individual self.

What theories of agency in psychology and connected disciplines often lack is the *contextualisation* of this phenomenon and its many forms of expression. Indeed, being agentic cannot be reduced to acting within a principle–agent relationship, to a pattern of neuronal connectivity, or an increased level of self-efficacy. We are agents in our daily lives in numerous ways and our action is always situated materially, symbolically, and socially. Even no action can be, and often is, a sign of agency. Added to this intrinsic diversity is the fact that theories of agency cannot ignore what Brockmeier (2009) calls *agentive discourses*. It is not only the case that human action bears the mark of agency but also this agency and the personal experience of being an agent are constructed through the way we think and talk about ourselves. Lay knowledge and experiential accounts of agency should not be disregarded since they participate in the phenomenon itself. Fundamentally, we are all seeking agentic forms of expression and we are also ready to attribute agency to others, including non-human elements (see Heider 1958). In the social situations we encounter, we are often driven by the need to explain the actions of self and others in causal terms or, in other words, to look for the responsible agents. In this process, we get to personify and invest with agency supra-individual entities (e.g. institutions or the state) and even nature (see Nash 2005). Moreover, we need also to account in our explanations for the possible difference of perspective between actors and observers and their sometimes conflicting interpretations of one and the same situation. And this state of affairs contributes to the proliferation of definitions I referred to above. However, surface diversity should not prevent us from observing some deep-level similarities across disciplines, as well as across science and common sense when it comes to theorising human agency; it is these similarities I move on to discuss next.

15.2 The Antinomies of Agency

At the core of the multiplicity of meanings attributed to agency in different disciplines stand a series of dichotomies that have historically structured our thinking about this phenomenon, both in science and lay understandings. This is not surprising considering that human thinking usually takes place in antinomies (Marková 2003) on the one hand and, on the other, considering the ongoing debates concerning

the nature and consequences of being an agent in the social world. Despite the fact that such debates tend to be over-productive in terms of theory, the conclusions they lead us to are not always illuminating. This is because the very act of establishing strict dichotomies can be misleading and frame the problem of agency in terms of one type or another when, in reality, the middle way or an integrative perspective would make more sense. As follows, I will focus on four key historical antinomies of agency which do not imply that these are the only polar lenses through which this phenomenon has been understood. In each case, I will reflect on the need to overcome dichotomous thinking and replace it with more relational forms of conceptualising agency (for a similar discussion of creativity, see Glăveanu 2013):

1. *Agency between complete freedom and the absence of freedom*. The importance of agency and the controversy over its extent relate closely to its association with the idea of freedom. For a long time in both philosophy and theology (see Browning Helsel; Chap. 10, this volume), the debate over agency revolved around 'free will', a concept that remains crucial for theories of action in philosophy and sociology, and for the study of ethics. The antonymic pair of freedom is determinism, a view that emphasises constraints over self-expression, predictability over chance, and dependence over independence. And this dependence is not only due to the influence of external social forces such as institutions or the state but is also organised internally through instincts and the unconscious. Deterministic theories are brought together by the conviction that there is no such thing as 'absolute freedom'. This claim is not hard to understand considering the fact that nobody lives in a material and social vacuum and, as part of larger biosocial systems, we are continuously influenced by elements or relations outside of us. And yet, the same systemic view (see Valsiner 2000, 2007) should make us sensitive to the fact that, just as we are constantly affected by environmental forces, we constantly exercise our influence over other people and the material world. This logic is in fact inscribed in the notion of 'relationship' seen as an interactive, bidirectional connection between person and context. As such, to argue for agency does not mean to argue for indeterminism (Williams 1992) but to consider freedom and constraints as two sides of the same coin. More than this, it means to overcome this false dichotomy by transforming it into *interdependence*. Just as complete freedom does not make sense in the absence of constraints, the absolute power of structures to determine our thinking and action fails to explain processes of personal and social transformation.
2. *Individual agency versus the agency of supra-individual agents*. If and when we accept the possibility of agency, the next important question becomes who exactly the 'agent' is. It is often the case that images of absolute freedom tend to be associated with a focus on individual actors and ground the capacity for agentic action within personal initiatives, particularly taken by 'creative' agents. This individualistic understanding of agency is contrasted by a more historical and sociological one, where the focus is on the state and its institutions credited as capable of creating and supporting social change. Considering the state a 'person' within international relations (Wight 2004) is challenged by other accounts

of social change that invite us to literally put people back into our understanding of nations (Thompson 2001). But the question remains in this latter case how we can explain macro-social change in terms of the action of single individuals. Burns and Dietz (1992) discuss this process by using evolutionary theory within a sociological framework. In order to do this, they define culture and institutions as systems of social rules and, consequently, envision evolution as the dynamic transformation of these rule systems. Individuals both produce and reproduce social norms. The evolutionary progress of social systems does not therefore aim towards any predetermined, final stage but is constantly shaped by complex contingencies and micro-level processes. In their turn, population-level phenomena regulate and select from the variations produced by individuals in a dynamic cycle of transformation coordinated at these different levels. In this case, it does not make sense to ask whether agency is specific for people, institutions or nations because, in reality, it does not reside in any but is produced by the relations between them.

3. *Agency as expressive versus agency as responsive*. Richard Shweder (1990), describing the connection between person and context, proposed a simple typology that is highly relevant for our discussion of agency. For him, this relation can be described as positive, whenever the intentionality of the world amplifies or supports that of the person and negative when, on the contrary, it opposes his or her intentions. Also, the relation is active when the person takes the initiative in creating or selecting aspects of the world most suitable for being acted upon and, reversely, passive when he or she ends up living in a world designed by others according to their own intentions. In other words, the person can be an agent or depend on the agency of others. What is interesting about this basic framework is that, in the spirit of what I have mentioned above, it focuses on the dual system of person and environment in conceptualising agentic action. In fact, psychology's own dilemma regarding human agency can be summarised in terms of expressiveness of the self, whenever individuals have the capacity to personalise their world, or in terms of responsiveness, being at the mercy of external stimulations and adapting to rather than transforming one's context. The idea of expressiveness, closely related to notions of authenticity and sincerity (see also Wilf 2011, for an anthropological perspective), portrays human beings as self-directed and capable of shaping the environment according to their aims and the emotional connections they establish with others. In contrast, the responsive approach describes humans as fundamentally passive beings, a view that is not far from the stimulus–response framework of behaviourism and, later on, cognitivism (see Bandura 2006). According to this perspective, we are mostly creatures of habit that get to learn structured ways of interacting with the environment based on previous experience. However, this account ignores the fact that habits themselves are open structures that allow people to flexibly (and agentically) navigate the world in ways that are simultaneously responsive to and expressive of personal life circumstances (see also the notion of habitus developed by Bourdieu 1990a, 1990b; Bourdieu and Wacquant 1992).

4. *Agency as originating in the brain versus the agency of the whole person*. One of the latest debates in the literature on agency originates in neuroscience and its recent quest for locating this phenomenon within the brain or, more broadly, to discovering the neurobiological correlates of agentic behaviour (for a discussion in the present volume, see Chaps. 4 and 5). From this perspective, agency is described in terms of processing information at different levels of the nervous system and generally associated with executive control. As such, it is considered to require the participation of different areas of the cerebral cortex such as the speech centres, face recognition areas, and the motor cortex (see Klemm; Chap. 4, this volume). What this localising effort is contrasted with is a broader understanding of agency as a property of the whole person. Agents cannot be reduced to their neurological processes but engage, at once, emotional, cognitive, and volitional elements, psychological phenomena that rely on but are of a different order than their biological substrate. This integrative image is specific for the scholarship of early psychologists and philosophers such as John Dewey (1934) for whom acting in the world necessarily requires the participation of our whole being, something that defines human experience. For Dewey, this totality goes beyond the isolated individual and includes human–environment relations. As I argued above, adopting a relational perspective on human agency helps us transcend dichotomies between the 'inside' and the 'outside', determinism and indeterminism, expressiveness and responsiveness, and, in this case, between the brain and the body. Contemporary cognitive science and philosophy of mind accounts are well aware of the dangers involved in trying to localise psychological functions in the brain, particularly phenomena as complex as agency, and are increasingly adopting more systemic approaches to the study of brain, body, and environment (see Hutchins 1995a, b; Gallagher 2013).

Whether agency is conceived as complete or constrained freedom, as the property of the individual or supra-individual actors, as self-expressiveness or mere responsiveness, as originating in the brain or in the totality of person (including person and environment), there are at least two underlying assumptions that permeate each of these antinomies. One concerns the (artificial) *separation between person and context* and the other portrays agency as *a unidirectional vector,* typically originating inside the person and directed towards the world; when agency is absent or limited, this unidirectionality takes the form of an imaginary vector pointing the other way around, from the environment to the individual. This kind of separation between person and world is at the root of several misconceptions fostered by dichotomic thinking since it supports the creation of boundaries, categories, and hierarchies. In order to transcend these divisions, what we need is a bidirectional concept of agency, one that considers this phenomenon socially and developmentally and places it within the relational space between person and his/her (social and material) context. A bidirectional type of thinking in this case invites us to theorise *co-agency* as a more suitable notion defining the co-participation of person and world in all types of actions we come to define as 'agentic'. Previously, other authors conceptualised this notion in relation to social networks (see discussions of distributed agency,

Yamazumi 2007; shared agency, Browning Helsel; Chap. 10, this volume) while in earlier work I focused mostly on the co-agency of humans and material objects (Glăveanu 2014). By articulating both and locating agency at the level of person–world relations, we are able to expand the meaning of agency itself and recognise it as a basic, perhaps the most pervasive and distinctive, feature of human existence. Agency understood not as a noun (a thing) but as a quality of our action is indeed omnipresent since both person and world are dynamic, open systems that shape each other through their relationship.

15.3 A Sociocultural Framework for 'Co-agency'

Before outlining a sociocultural framework for the notion of co-agency, I will begin this section with two examples drawn from everyday life, one represented by a recent episode of shopping for lunch with a friend in London and the other taken from my extensive fieldwork with Easter egg decorators in a village in northern Romania. These two mundane situations bring to the fore several important issues for our discussion of agency such as the interplay between choice and constraints, the role of habits and tradition in structuring action, and the possibility of change within scripted interactions. Both shopping and decoration activities are, to a different extent, structured by social and material constraints and yet their outcomes are often different from what we expect or intend to achieve. This is a common situation, for example, when entering a supermarket, even when a shopping list was prepared from before, and it is certainly the case of decorated eggs whose colours and shapes often surprise their makers. Are these to be considered signs of agency? Or should we interpret them as the result of circumstance and hazard? As I come to show within this section, agency typically builds on chance and circumstance but always does so in a reflective manner. Moreover, this process is not led by the person alone but shared with other people and even with the artefacts that surround us and 'respond' to our action.

Going shopping for food is certainly a widespread activity and one that builds on a number of culturally established scripts (Stoltman et al. 1989). The organisation of the supermarket space supports this normativity by regulating the movement of customers. There are certain points of entry and certain points of exit (and indeed confusing them might call the attention of the staff). Baskets are conveniently placed near the entries and different corridors define the path one can take inside the supermarket. Moreover, food and other products are displayed according to a certain logic within the shop (although this logic often escapes hurried clients), and grouping goods in different pre-established categories is the norm. If the customer is a regular of the place, the shopping process is largely facilitated by this (relatively stable) organisation and made more efficient. However, at least once in a while, a change in display or an advertisement for a promotion might disturb our routinised ways of navigating the space of the supermarket. First-time customers will almost always need to pay attention to environmental cues in order to understand the place

and find their way within it according to what they wish to purchase. But in most cases, again, the general normativity of shopping permeates this activity and facilitates its success.

The episode I will briefly describe here took place recently while buying a few items for lunch together with a friend in London. On that day, my friend and I were walking past a well-known supermarket in the UK when we remembered that we wanted to buy some food for my lunch (as we had decided to eat at home). Entering the shop with the rather vaguely defined intention of buying some already cooked food, I started by looking for the section that had such items. Since the UK food culture encourages one to buy takeaway food for lunch such as sandwiches, salads, or pasta, I was spoiled for choice and this, I noticed, was not necessarily the best thing for my decision-making process. While having options and the possibility of (freely) choosing between them is at the core of what we understand by agency, too much choice and too much freedom can lead to indecisiveness and, in some cases, to abandoning the situation. Furthermore, how free are we to choose in the supermarket? One might argue that, even without realising, our agency is constrained, manipulated even, by the display of the items and marketing strategies, including pricing (e.g. Fiore et al. 2000). On the other hand, I have some relatively well-defined preferences, and it would be hard to be manipulated into changing them on the spot (of course, deterministically minded people can inquire further into how I came to form these preferences but we will leave this aside for now). And indeed, after exchanging a few items, I saw something I immediately knew I preferred (if you are curious, it was a box of paella) and proceeded to buying it after inspecting the label and content (through the transparent packaging). Interestingly, during this process, I told my friend who is a vegetarian that I wanted to buy a vegetarian dish in order not to bring meat, although packed, into her fridge. Unfortunately, I ended up doing the opposite, but only because I knew she would not mind me choosing the food I wanted.

The second example I want to bring to this discussion takes us to a completely different space than that of a big supermarket in a busy metropolis. In contrast, this context is that of a quiet village situated between mountains in the north of Romania, close to the border with Ukraine. Ciocăneşti might be a small rural space but is well known in the country for its active community of egg decorators, a tradition with ancient roots in Romania (Gorovei 2001; Zahacinschi and Zahacinschi 1992). Using an old system of decoration techniques and motifs, this practice typically requires the artisan to work with warm wax on the surface of the egg and successively immerse it in colour, particularly yellow, red, and black (for details about the process, see Glăveanu 2012). The act of decoration is, as such, deeply embedded within the material and symbolic world of the artisan and his or her community. Its outcomes reflect the taste and aesthetics of different regions within Romania and it is often the case that eggs decorated in one village differ quite substantially from the designs used in another village close by. On the whole, though, in the north of the country, geometrical decoration is preferred despite the fact that recent decades brought about many changes in the life of this craft. For example, it is rare nowadays to use natural colour pigments and more intricate patterns are preferred to

simpler, traditional motifs; in addition, eggs with wax in relief or figurative decoration are becoming more and more common. These changes have been driven by an ever-growing market, both national and international, for the distribution of these artefacts.

As part of my fieldwork, I interviewed and observed decorators while working in an effort to unpack microgenetic creative processes within a traditional craft. As part of this ethnographic study, I closely documented the activity of known decorators in Ciocăneşti through photographs and videos (collected with the help of the subjective camera, for details see Glăveanu and Lahlou 2012) at the home of the artisans, at the Museum of the Decorated Egg, and at the national Easter fair. The series of pictures depicting the first stage of decoration (see Fig. 15.1) were taken at the national fair and present the work of Valerica Juşcă, one of the participants at the event. As these pictures show, Valerica started by segmenting the egg with the help of a pencil and not directly with wax. This is a widespread procedure that helps decorators create reference points before drawing the motifs in wax. It is also an additional safety measure since depicting ornaments on an object the shape and size of the egg (in this case hen eggs, the only ones admitted at the decorators' contest) is challenging and any mistake made in wax cannot be repaired later on (as wax sticks to the surface of the egg and, even when erased, makes colour 'catch' less well on that particular surface). Being an experienced artisan, Valerica drew double and not simple lines, effectively segmenting the egg into two main areas: the extremities (top and bottom) and a larger, middle belt. This kind of segmentation is common but there are also other possibilities and the decision is largely influenced by the shape of the egg, one of the many ways in which the materiality of the artefact participates in the agency of the decorator. On the top part, Valerica decided to alternate star rays and hooks, both simple and traditional ornaments that can be combined and re-combined in different ways and associated with many other elements (in fact, they are themselves taken from more elaborate motifs such as the star with eight rays, a symbol of perfection and femininity). The belt part was segmented into four main quadrants and, just as with the top, symmetry guided decoration work. Valerica chose in one of these quadrants (repeated later on the other side of the egg) to continue depicting star rays and hooks, something perhaps prompted by the design of the top. However, due to the different disposition of the space, she also included a cross in the middle, a Christian symbol designating the religious significance of these artefacts. In the other quadrant (also repeated later on the other side), she opted for a traditional pyramid-like shape on the sides of diagonal lines. This shape is often found on ornamental houses in the community and testifies to the ways in which folk art ornaments communicate between different media, from eggs to houses, gates, and clothes. Finally, the triangular shapes ending in a dot remind of the ends of towels and tablecloths (see also the lines drawn around the triangular shape of the star rays) and, once more, suggest a close dialogue between this craft and the material and cultural world it belongs to.

These two examples, of shopping for lunch and decorating an Easter egg, perfectly illustrate the dynamic negotiation, within agentic action, between freedom and constraints, between personal initiative and the intentionality of other people

Fig. 15.1 Egg decoration in Ciocăneşti (Valerica Juşcă)

Fig. 15.1 (continued)

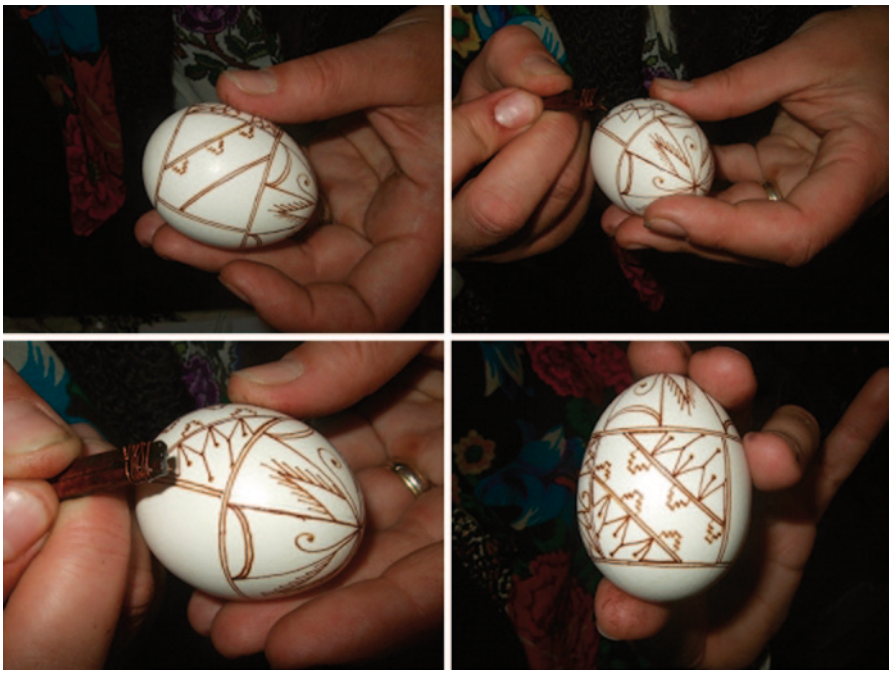

Fig. 15.1 (continued)

and objects, between the possibility for change and the normativity of tradition and social scripts. Navigating the space of the supermarket and ornamenting the space of the eggshell are opportunities for agency not because they lack any determinism but precisely *because* they are highly constrained, materially, symbolically, and socially. As an agentic individual, I am very much used to entering shops, including the branch I visited during my trip to London and, similarly, Valerica has been decorating eggs for years using motifs such as the star or the shepherd's hook and will probably continue to do so in the future. And yet, our actions never lead us to the same outcomes. I rarely, if ever, had that type of paella for lunch, and arguably the egg produced by Valerica is absolutely unique. The variability inscribed in both person and context, because of the open and developmental nature of both, is one of the main engines of our agency and, as I argued elsewhere, of creativity (see Glăveanu and Gillespie 2014). What is though specific for agency? I propose as follows that there are at least three key elements present in both situations described above and potentially found in a variety of human activities. These are *intentionality, (re)action,* and *reflexivity*. The framework I develop and discuss below defines co-agency as the *cyclical coordination* between these processes, all placed within a context characterised by numerous social and material constraints (see Fig. 15.2). Within this cyclical interrelation, the intentions, actions, and reflections that make up our agency are constantly distributed between person and his/her environment, thus emphasising the fact that agency itself is best understood as 'co-agency'.

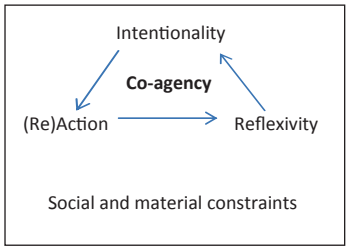

Fig. 15.2 A framework of co-agency

Proposing a cyclical framework of co-agency means to consider this phenomenon as a (relational) process rather than a thing or static quality. It also means that there is no particular start or end point within this dynamic and that, indeed, the three 'moments' of agency are fundamentally interrelated in any concrete situation, they overlap and feed into one another. However, for analytical purposes, we can begin by discussing the notion of *intentionality* in the model above. From the start, it is important to note that I refer here to intentionality and not intention since I consider the former to be a broader orientation of the person towards the world, a general sense of directedness (for a philosophical account, see Brentano 1874). This stands in contrast to usual goal-directed accounts of action (Chranach 1982) in which the intention to achieve a particular result and the representation of this result are given priority. While I do not fully disagree with this approach, I do believe that by reducing intentionality to specific intentions, we are not only missing the richness of this concept but more easily fall prey to descriptions of action that separate intentions from the process of accomplishing them. For example, in the creativity literature, there tends to be a general consensus that one's work is creative only if it comes out of an intentional process in which the creator is largely aware of his or her creative goals (Weisberg 1993). However, what happens in reality, and what is clearly the case for Valerica in her practice of egg decoration, is that intentions or goals are formulated and reformulated *in the very process of acting on them*. Intentionality is indeed not an antecedent of action, separate from it and determining its course. As part of the co-constructive process of agency, our intentions originate from the interaction between person and context and not solely from 'inside' the individual. Consider the shopping episode briefly described above. Entering the supermarket, I did not have the specific intention to buy a certain type of food and the decision process was shaped by an active exchange between my preferences, the dialogue with my friend, and the choices available (including their material properties as reflected by packaging, pricing, etc.).

As such, in practice, the distinction between intentionality and action is blurred. To understand them better in their unity, we need to consider next the notion of *(re)action* as one that captures the simultaneously active and reactive aspects of our relation to the world and both its human and non-human agents (see also Latour 2005). To act in and on the world involves continuous *action–perception loops* within which the person creates a certain change in the environment and this creates at the same time a change within the actor himself or herself. We not only act but also perceive the outcomes of our action and are influenced by them. This is the case

of the artisan decorating eggs whenever a movement within the decoration process can continue or destroy the work (see the notion of workmanship of risk, Pye 1968), something that prompts the person to find practical solutions of an improvisational nature. It is what Dewey (1934) famously referred to as the interconnection between doing and undergoing, between directing one's action towards the world and then incorporating its reaction to it. (Re)Actions are crucial for agency because, in essence, they are the ground for both the making and unmaking of intentionality and reflexivity. In this sense, once more, to think of agency as a property of the person's doing is limited. This unidirectionality is complemented and, indeed, shaped by the resistance of the environment towards one's doing. This in-between-ness of action is reflected also in the fact that what it typically does is exploit the affordances of objects and social relations. And affordances are themselves relational properties that reunite the possibilities of the person with those of the material and social world (for a broader discussion, see Gibson 1968; Costall 1995). If I was able to enter the supermarket, walk down the aisles, and find the food I wanted, it was because the environment was structured in intentional ways (see Shweder 1990) and I made use of existing affordances (of doors, shopping baskets, packaged and ready-cooked food, money, etc.). The intentionality of other people—in this case those who produce and sell the goods—is inscribed in the way the material world is organised and this both facilitates and constrains my agentic (re)action within the environment.

And it is sometimes the case that constraints are very visible and end up blocking the path of action and making the person pause and reflect on the situation. This happens particularly when confronted with novelty, with the unexpected, or with obstacles. In the pragmatist tradition, such moments are considered to be at the origin of thinking processes (Mead 1964) and, in another strand of literature, of creativity (Torrance 1988). I have discussed elsewhere why connecting conscious, creative action with the presence of obstacles might distort our view of everyday activities, habits, and routines (see Glăveanu 2013). However, for the purpose of this framework, my definition of *reflexivity* will draw more on dialogicality and perspective taking[2] (see Gillespie 2006) and formulate it as the capacity to adopt a new perspective on one's action and its outcomes. In this sense, being reflective is a deeply social act and it also has the potential to shape our intentionality and future course of action. Reflexivity was connected by Dewey with undergoing, the moment within action when the person becomes aware of his or her goals. As such, there is once more a blurred line between (re)action and reflexivity within the framework proposed above and this is demonstrated also in my supermarket example by the fact that, within the process of buying lunch, I started considering different types of food from the perspective of my vegetarian friend. This act of reflexivity was not taking place inside my head but expressed in both action (taking or putting back items, etc.) and communication (letting my friend know about my intention, asking if she would mind me buying finally the paella, etc.). Equally, there is reflexivity in the process of egg decoration although it might not be always expressed verbally since most of the times this work is done alone or in silence.

[2] I am grateful to Constance de Saint-Laurent for discussions of this notion that centre on the notion of perspective.

And yet, filming this practice, I could notice how both expert and novice decorators regularly turn the egg around and look at the motif at intervals of a few seconds (see Glăveanu and Lahlou 2012). Asked what this move accomplished, I was told it involves taking stock of what has been done and reflecting on future steps. Within the framework of co-agency proposed here, this is a crucial phase in the process, in which decorators take the perspective of the audience on their own work (e.g. of peers, of future customers).

Finally, we can also note that, in Fig. 15.2, co-agency is depicted as embedded within *social and material constraints* which both limit and allow what agents can actually do. The absence of any constraint is blocking (think about the instruction 'be creative!' given without any other details) and so is the presence of too many restrictions (because it can be demotivating). What the 'optimum' is in this case depends on both person and context of action and the dynamic articulation and co-evolution of both. In fact, one of the key characteristics of the sociocultural model put forward in this section resides in its emphasis on connectivity and temporality. As a unitary system, agent and environment shape each other within the cyclical movement between intentionality, (re)action, and reflexivity. And this broader and contextual understanding sets this model also apart from other conceptualisations, much more individual in focus, including some presented in this book (see Klemm's (Chap. 4, this volume) discussion of the following chain of stages: intend, remember, value, decide, prepare/plan, and act). However, other authors did consider more of the elements described here. For instance, Bandura (2006, pp. 164–165) presents an interesting commentary on what he calls the four core properties of human agency: intentionality, forethought, self-reactiveness, and self-reflectiveness. On the surface, these terms he uses are strikingly similar to those in Fig. 15.2, and yet a closer analysis reveals that his conception of intentionality refers to formulating action plans and strategies (in which forethought designates cognitive representations of these future plans and goals), reactiveness means self-regulation, and reflexivity is represented by functional self-awareness and metacognitive capacities. For some readers, particularly those more inclined towards adopting a cognitive approach, such a vocabulary will sound reassuringly familiar. However, in my view, it is too individual centred and intra-psychological to account for changes in the world and not only in the self, a key concern for any theorist or researcher of co-agency.

15.4 Co-agency in a Broader Framework

In this chapter, my main aim was to introduce and develop the concept of co-agency and, in doing so, I tried first to position this conception within the landscape of various disciplinary perspectives on the topic, many of them reflecting a long tradition of theorising what is specific for human beings and for human society. Many of these perspectives can be organised along the lines of simple antinomies, an exercise that is not always fruitful as it tends to radicalise our view of agency. Trying to add to the existing literature and go beyond dichotomic thinking, I proposed here a basic (and triadic) sociocultural framework of co-agency describing *shared or*

distributed processes between people and their (social and material) environment, processes that contribute to the cyclical relation within co-agency between intentionality, (re)action, and reflexivity. While I am the first to acknowledge the fact that such divisions are primarily analytical, I do believe discussions of agency necessarily have to deal with the issue of intentionality, with the dynamic between doing and undergoing, and the capacity to reflect on the outcomes of one's action from the position of other people. One benefit of looking at such primary processes is that we are able to overcome the idea, rooted in both scholarly and lay conceptions, that agency is specific for a limited number of individuals who are capable, unidirectionally, to shape culture and society and position themselves outside the deterministic forces of conformity and routine. In contrast, co-agency considers as its basic unit the *agent in context*. It does not deny the role of the individual person but conceives of this person as a thoroughly sociocultural agent, one that does not stand apart from his or her environment but acts from within it. Human agency becomes thus a widely distributed process not to be found only in the hands of a few but in the everyday life of all. This universality of co-agency largely resides in the basic fact that, through our capacity to symbolically act on the world, we are capable of detaching ourselves from the here and now. We are thus not trapped by our immediate reality but use tools and signs (Vygotsky 1997) to respond to and also transform our environment. In this sense, symbolic activity as expressed in meaning-making represents one of the most distinctive signs of human agency (Brockmeier 2009).

While acknowledging such primary processes that focus our attention on the psychological, we must also consider the ways in which co-agency operates between people and how it can be understood as an interpersonal rather than individual phenomenon. The theoretical framework advanced in this chapter captures the dialogue between agents and their context by stressing the co-constructed nature of intentionality, the dynamic between acting and being acted upon, and finally, the social positions embedded within moments of reflexivity. Chapter 10 points to the fact that agency is not the only 'good' or 'virtue' significant for human life and mentions, among others, community and connection, acting in a trustworthy manner, and being capable of self-sacrifice. I would argue, from my perspective, that co-agency necessarily builds on community and connection and should always be considered, theoretically and methodologically, in relation to these constructs. Bandura (1982, 2000) himself, when linking agency and self-efficacy, referred consistently to both individual and collective forms of expression for both (the latter relating to the shared beliefs we hold about the power of collective action). But Helsel's arguments make us consider not only the community behind agency but also the ways in which shared agency itself contributes to the existence of the community. It raises issues of *personal and social responsibility* for those who are denied or lack agency in particular contexts. Indeed, no discussion of co-agency is complete without recognising *relations of power* between different agents (human but also non-human, particularly supra-individual) and the way they shape our intentionality, capacity for action, or guide our reflective efforts.

If the sociocultural framework elaborated in this chapter offered an answer to the fundamental question 'do humans have agency?' (by arguing that yes, co-agency is

widely distributed and represents a marker of our everyday existence and experience of the world), a focus on inequality, domination, and power structures invites us to reformulate this interrogation into the more urgent and pragmatic question of 'how do we *foster* co-agency in the case of individuals and communities?'. This type of reframing forces us to understand that, although our actions carry with them the marker of co-agency, this inherent potential is not always realised to its fullest; there are indeed many people and communities in the world who, in one way or another, are prevented from becoming agents in their own right and developing their own forms of intentionality and action or building forms of reflexivity that would allow them to imagine and construct a better future for themselves. Without being void of any agency, we often have the experience of not being the agents of our own development within concrete situations (e.g. think about the experience of many students in formal education settings or that of people living in totalitarian regimes). But, precisely because agency defines bidirectional, evolving relationships, there is always the *possibility of resistance* and we have numerous examples throughout the world of how individuals and communities fight against hegemonic discourses, colonising practices, and dehumanising forms of discrimination (Howarth 2004; Jovchelovitch 2014; Glăveanu and Sierra, under review). The question for theorists and practitioners alike is how to support their efforts and how to contribute to empowering the marginalised and the oppressed? Although there are many factors to take into account and each case needs to be considered separately, we might draw some inspiration from the sociocultural framework of co-agency and consider, at a practical level, what kinds of intentionality are being constructed with or for people, how action is met by obstacles and, most importantly, how this experience can lead to reflexivity on the part of all those involved.

In the end, co-agency is never a given but a constructive, relational process, always in the making within the interaction between individuals and groups. As a shared phenomenon, people 'take part' or 'participate' not only in their own agency but also the agency of others. Today's world and its increased connectivity through, among others, the Internet and new media offer both the paradigmatic example of co-agency and represent its necessary condition. We live in a time and age in which opportunities for communication, exchange, and cooperation abound. And yet, this is also a world in which the digital revolution creates the conditions for mass surveillance and the development of more and more sophisticated ways of limiting the agency of people often without them even realising it. How far is this reality from the experience of Winston Smith in Orwell's 1984? Clearly, there are visible differences when it comes to the level and means of coercion used. In Smith's case, the aim of his persecutors was to destroy any intentionality or personal initiative, reduce possibilities for action and thought, and, ultimately, eliminate reflexivity or the capacity to question authority (Big Brother). The reality of our lives and interactions with others is however more complex. But if we do take co-agency seriously, we need to consider carefully both the person and the material and social system it belongs to and observe the development or, on the contrary, the reduction of agency within their interrelation. We also have to be aware of the ethical nature of this relationship and start theorising *moral co-agency* as the basis for better, more fair, and inclusive societies.

References

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175–1184.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9*(3), 75–78.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*(2), 164–180.

Bourdieu, P. (1990a). *In other words: Essays toward a reflexive sociology*. Stanford: Stanford University Press.

Bourdieu, P. (1990b). *The logic of practice*. Stanford: Stanford University Press.

Bourdieu, P., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Chicago: University of Chicago Press.

Brentano, F. (1874). *Psychology from an empirical standpoint*, (1973 edition). London: Routledge.

Brockmeier, J. (2009). Reaching for meaning: Human agency and the narrative imagination. *Theory & Psychology, 19*(2), 213–233.

Burns, T. R., & Dietz, T. (1992). Cultural evolution: Social rule systems, selection and human agency. *International Sociology*, September, *7*(3), 259–283.

Cheliotis, L. K. (2006). How iron is the iron cage of new penology? The role of human agency in the implementation of criminal justice policy. *Punishment & Society, 8*(3), 313–340.

Costall, A. (1995). Socializing affordances. *Theory & Psychology, 5,* 467–481.

Cranach, M. von. (1982). The psychological study of goal-directed action: Basic issues. In M. von Cranach & R. Harré (Eds.), *The analysis of action* (pp. 35–73). Cambridge: Cambridge University Press.

Dewey, J. (1934). *Art as experience*. New York: Penguin.

Fiore, A. M., Yah, X., & Yoh, E. (2000). Effects of a product display and environmental fragrancing on approach responses and pleasurable experiences. *Psychology & Marketing, 17*(1), 27–54.

Friedman, B., & Kahn, P. H. Jr. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software, 17*(1), 7–14.

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research, 25/26,* 4–12.

Gell, A. (1998). *Art and agency: An anthropological theory*. Oxford: Clarendon.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale: Lawrence Erlbaum.

Gillespie, A. (2006). Games and the development of perspective taking. *Human Development, 49,* 87–92.

Glăveanu, V. P. (2010). Principles for a cultural psychology of creativity. *Culture & Psychology, 16*(2), 147–163.

Glăveanu, V. P. (2012). Creativity and folk art: A study of creative action in traditional craft. *Psychology of Aesthetics, Creativity, and the Arts, 7*(2), 140–154.

Glăveanu, V. P. (2013). From dichotomous to relational thinking in the psychology of creativity: A review of great debates. *Creativity and Leisure: An Intercultural and Cross-disciplinary Journal, 1*(2), 83–96.

Glăveanu, V. P. (2014). *Distributed creativity: Thinking outside the box of the creative individual*. Cham: Springer.

Glăveanu, V. P., & Gillespie, A. (2014). Creativity out of difference: Theorising semiotic, social and temporal gaps. In V. P. Glăveanu, A. Gillespie, & J. Valsiner (Eds.) *Rethinking creativity: Contributions from cultural psychology* (pp. 1–15). London: Routledge.

Glăveanu, V. P., & Lahlou, S. (2012). 'Through the creator's eyes': Using the subjective camera to study craft creativity. *Creativity Research Journal, 24*(2–3), 152–162.

Glăveanu, V. P., & Sierra, Z. (under review). Creativity and epistemologies of the south. *Culture & Psychology*

Gorovei, A. (2001). *Ouăle de Paşte. Studiu de folclor* (2nd ed.). Bucureşti: Paideia.

Gough, N. (1999). Rethinking the subject: (De)constructing human agency in environmental education research. *Environmental Education Research, 5*(1), 35–48.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
Hollis, M., & Smith, S. (1994). Two stories about structure and agency. *Review of International Studies, 20*(3), 241–251.
Howarth, C. (2004). Re-presentation and resistance in the context of school exclusion: Reasons to be critical. *Journal of Community & Applied Social Psychology, 14*(5), 356–377.
Hutchins, E. (1995a). *Cognition in the wild*. Cambridge: MIT.
Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science, 19,* 265–288.
Jovchelovitch, S. (2014). The creativity of the social: Imagination, development and social change in Rio de Janeiro's favelas. In V. P. Glăveanu, A. Gillespie, & J. Valsiner (Eds.), *Rethinking creativity: Contributions from cultural psychology* (pp. 76–92). London: Routledge.
Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press.
Lavazza, A., & De Caro, M. (2010). Not so fast. On some bold neuroscientific claims concerning human agency. *Neuroethics, 3*(1), 23–41.
Libet, B. (2002). Do we have free will? In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 551–564). Oxford: Oxford University Press.
Lubart, T. (2003). *Psychologie de la créativité*. Paris: Armand Colin.
Marková, I. (2003). *Dialogicality and social representations: The dynamics of mind*. Cambridge: Cambridge University Press.
Mead, G. H. (1964). *Selected writings: George Herbert Mead*. A. J. Reck (Ed.) Chicago: University of Chicago Press.
Nash, L. (2005). The agency of nature or the nature of agency? *Environmental History, 10*(1), 67–69.
Pile, S. (1993). Human agency and human geography revisited: A critique of 'new models' of the self. *Transactions of the Institute of British Geographers*, New Series*, 18*(1), 122–139.
Pye, D. (1968). *The nature and art of workmanship*. London: The Herbert Press.
Reed, M. I. (1988). The problem of human agency in organizational analysis. *Organization Studies, 9*(1), 33–46.
Runco, M. (2007). *Creativity. Theories and themes: Research, development, and practice*. Burlington: Elsevier.
Shapiro, S. P. (2005). Agency theory. *Annual Review of Sociology, 31*, 263–284.
Shweder, R. (1990). Cultural psychology—What is it? In J. Stigler, R. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 1–43). Cambridge: Cambridge University Press.
Stoltman, J. J., Tapp, S. R., & Lapidus, R. S. (1989). An examination of shopping scripts. In T. K. Srull (Ed.), *Advances in consumer research* (Vol. 16, pp. 384–391). Provo: Association for Consumer Research.
Thompson, A. (2001). Nations, national identities and human agency: Putting people back into nations. *The Sociological Review, 49*(1), 18–32.
Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 43–75). Cambridge: Cambridge University Press.
Valsiner, J. (2000). *Culture and human development: An introduction*. London: Sage.
Valsiner, J. (2007). *Culture in minds and societies*. New Delhi: Sage.
Vygotsky, L. S. (1997). The history of the development of higher mental functions. In R. W. Rieber (Ed.), *The collected works of L.S. Vygotsky* (Vol. IV, pp. 1–251). New York: Plenum.
Weisberg, R. (1993). *Creativity: Beyond the myth of the genius*. New York: W. H. Freeman.
Wilf, E. (2011). Sincerity versus self-expression: Modern creative agency and the materiality of semiotic forms. *Cultural Anthropology, 26*(3), 462–484.
Wight, C. (2004). State agency: Social action without human activity? *Review of International Studies, 30*(2), 269–280.
Williams, R. N. (1992). The human context of agency. *American Psychologist, 47*(6), 725–760.
Yamazumi, K. (2007). Human agency and educational research: A new problem in activity theory. *Actio: An International Journal of Human Activity Theory, 1*, 19–39.
Zahacinschi, M., & Zahacinschi, N. (1992). *Ouăle de Paşti la români*. Bucureşti: Editura Sport-Turism.

Chapter 16
The Explanatory Power of Agency

Sven Hroar Klempe

Agency is first of all interesting in terms of its explanatory power. However, there are several uncertainties and ambiguities related not only to agency but also to explanations, and not least their more or less actual power to explain. In his classic pamphlet on explanations, Toulmin (1963) demonstrated that no explanations are able to provide adequate predictions, but their qualities are rather related to their capacity to make sense of the connections between the observed facts.

The ambiguity of the power of explanations is envisaged in the immediate understanding of explanatory power. A theory that explains an abundance of events is normally to be preferred, whereas a theory that is rather limited and explains just a few events is rather understood as being weak. Yet psychoanalysis and Darwinism exemplify theories with strong explanatory power, but that is exactly also why they are criticized. They explain too much and it is hard to come up with competing explanations. Hence, this is why Karl Popper turned these preferences upside down and replaced confirmation, which is related to theories with strong explanatory power, with falsification, which is rather related to theories with weak explanatory power (Popper 1980, p. 403).

16.1 To Approach Meaning

The introduction of falsification as a criterion for science in general was refuted by Charles Sanders Peirce before Popper even had formulated it (Peirce 1878/1986). Peirce pledged rather for a kind of grounded belief as the criterion for science. This highlights Toulmin's point, specifically that explanations might be regarded as a kind of justified meaningful understanding of an event. If so, the crucial factor in science is first of all the aspect of justifying and how the theory is grounded. How-

S. H. Klempe (✉)
Norwegian University of Science and Technology (NTNU), Trondheim, Norway
e-mail: Hroar.Klempe@svt.ntnu.no

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9_16

ever, this does not necessarily bring us to grounded theory, although the original aim of Glaser and Strauss was exactly to make theories grounded and justified (Glaser and Strauss 1967). Grounded theories are not just based on empirical aspects, but even as much on appropriate theoretical reasoning and consistency. This formed the basis for all scientific activities in the wake of Aristotle, which lasted for more than 1500 years in terms axiomatic method and demonstrative sciences. Yet along with the entrance of the modernity around 1600, the Aristotelian conception of axioms and demonstrative science became challenged.

A focus on agency may also be regarded from this perspective. It represents a typical modern interest, and is closely related to autonomy, secularization, democratisation and individualism. It is about the human being defined as an independent entity. Yet this is the challenging paradox at the same time, specifically that the independent human being should apparently be able to line up the depending variables by which its acts are to be explained. This is a challenge, not only for explanations but also for how to approach and achieve an explanation. Thus, there are a lot of approaches for explaining human agencies because the angles from which they are to be explained may differ. One of the oldest is the hermeneutic approach, which is not just one, but represents at least three different approaches due to the historical development. First, there is the text in context (theological exegesis from the fifteenth century), then there is the text and historical context (Dilthey 1996), and the third is the reader as an existential premise for both the text and the historical context (Gadamer 2004). Those three versions of hermeneutics illustrate how agency may be embedded with both paradoxes and huge explanatory challenges.

16.2 Three Explanations

To clean up this confusion, one may look at different types of explanations. Although it is normal to bring in the four types formulated by Aristotle—the material, the formal, the final and the efficient causes—these can either be seen as old fashioned, or even fundamentally misunderstood (Heidegger 1973). Those types of explanations that are applied in science today are rather (1) the efficient cause, (2) the intentional cause and (3) the functional cause (Elster 1979). The way the efficient cause is understood today is quite different from how Aristotle defined it. Today, it refers to the blind, physical forces that interact in nature. Intentional causes are rather characterized by the opposite. An event may have an intentional cause when it is to be regarded as a result of an individual's acts provided by conscious purposes. Thus, the event is a result of mental states like desires or beliefs. The functional explanation, on the other hand, represents a more open type of explanation, in which the relationship and, more specifically, the direction between the cause and the effect, is more open, and hard to define.

These three groups of causes should apparently make it easy to see how agency could be applied as an explanatory term in psychology. Agency is first of all related to intentional causes although this is not so obvious as it may look like. Intentional

causes are closely related to and even presuppose conscious intentions behind the acts. This requirement goes back to Kant who regarded unconscious intentions as a contradiction in terms. And he was of course right as long as we talk about clear intentions as the cause. However, to reduce human acts to clear intentions implies that the individual should have full control of even the smallest bodily movement, like the slightest wink of the eye. Psychology is full of examples demonstrating that each one of us do not have that kind of control. All the aspects of volition, desires, values and even automatized behaviour bring in so many additional factors, which leave intentions blurry and concealed by an abundance of reasons that may lie behind an act. Hence, the term agency is a convenient term to apply because it does not presuppose a clear intention, but points in different directions.

This leads immediately to the third type of explanations, which is the functional. This might be defined in many ways, among which a loose version could be 'a situation in which the effect alone is regarded as the cause' (Klempe 2014, p. 72). Although this type of causality is very close to the Aristotelian final cause because the effect is always to be regarded as the same as the result, it is still applied. Yet the justification of this type of explanation is first of all given by its ability to create meaning out of a situation. And meaning is made when we can see connections, no matter in which directions the connections are supposed to go. This is also why statistical proofs are so desirable, not because they reveal causality, but rather because they are able to create unspecified connections in terms of correlations. If there were any arguments that led to connections between two or more events, those would quite likely be supposed as sufficient for making the statements meaningful.

16.3 The Explanatory Power of Coagency

The individual's efforts in making meaning out of a situation can provide an abundance of psychological explanations. Yet these efforts are first of all related to the ultimate existential dilemma in an individual's life, which is a result of the fundamental conflict between experienced coherence on the one hand and experienced discontinuity on the other. As long as those two extremes are not or maybe even cannot be united, life as such appears as diverse and multifarious. The attribution theories in social psychology have told us a lot about how the situation decides how we make connections and try to build up an understanding based on not only our own understanding of the situation but also our perceptions of how other perceive and understand the same situation. A certain situation is rather a fretwork of intentions, opinions and judgments, communicated on different levels at the same time, which challenges all our ideas about causality related to agency.

Yet if this forms the picture, our scientific intuitions force us to take just one step at the time. One of these is to move carefully from agency to coagency (Glăveanu; Chap. 15). Although this is just one step, it represents a radical change. This is the qualitative step from unity to manifold, where the difference between the two steps coincides with the above-sketched fundamental existential conflict between

coherence and discontinuity, or the experienced differences between attachment and separation. In other words, to go from agency to coagency is the same as breaking down the harmonious relationship between events and rather open up for more clashing dissonances. However, that was exactly what social psychology opened up for already in the 1950s, by focusing on concordances and conflicts in actions and interactions among people (Festinger 1957). Yet still the final word was not said when it comes to making meaning out of social conflicts in terms of being rational and understandable.

In this respect, the idea of cyclical coordination between the three factors: intentionality, (re)action and reflexivity (Glăveanu; Chap. 15) mitigates in a very interesting way the size of the gap between agency and coagency. First of all, it brings in the rational aspect of intentionality by being a cause for actions. However, these are not governed by complete control, because the actions are moderated by a kind of more or less unconscious awareness of reactions from other members of the social group. Moreover, the aspect of reflexivity is to be regarded as a dynamic process. To reflect is not to have an overall and unified thought, but to achieve one if possible. This implies that reflections reflect the deep and probably irreconcilable conflict between the one and the other—between the one and the many. So, if we then unite the three factors: intentionality, (re)action and action, the aimed unity is provided by an unspecified original intention, the multiplicity is provided by actions and reactions, and the endeavour of unifying the multiplicity is provided by reflections.

Thus, the explanatory power of the term coagency must be said to be strong. It is probably stronger than just agency because it includes different agents and says something about how they coordinate their acts. It is probably also more precise for exactly the same reasons. On this basis, it is possible to formulate some axioms that reflect the benefits from the term ‘coagency’. Hence, the axiom of coagency:

> Human agency can only be explained in terms of the aspect of cooperation among agents by means of intentions, actions and reflections.

This implies that the explanatory power of agency is primarily provided by the term ‘coagency’.

16.4 Transdisciplinarity and Explanatory Power

An attempt at understanding agency by bringing in coagents is just one step, but most likely the first step to take. The next step to complete this achievement is probably to include a transdisciplinary approach (Uher; Chap. 13). Transdisciplinarity might be defined in terms of ‘a different manner of seeing the world, more systemic and more holistic’ (Max-Neef 2005, p. 15). This is always challenging a discipline because their protectors and spokesmen will always argue for their discipline’s independency and unique scientific traditions.

This is also true for psychology, although it has a quite complex history, and this history is not necessarily so very old either. The age is very much dependent on

where we put the point zero. The most widespread alternative is related to Wundt's laboratory in Leipzig in 1879. Although this is the year that is referred to by most of the textbooks, it is probably a date even Wilhelm Wundt himself would have had objections to. He regarded experimental psychology just as a subordinated 'Hilfsmittel' in psychology since it did not include thinking and language (Wundt 1920). The use of the term 'psychology' is also quite new in the sense that the term can only be traced back to the sixteenth century (Mengal 2005; Vidal 2011). Although psychology in the past may have been something completely different from how it is conceptualized now, there are some aspects that can tie them together though, and one of those is the aspect of transdisciplinarity.

If we say that there is a difference between 'transdisciplinarity' and 'interdisciplinarity', one of the differences should be that the latter presupposes borders between the different disciplines, whereas the former does not. If this is true, then we have to take a step back in the history to see if psychology transcends the narrow frameworks that defines it today. And this is exactly what we may find. The conception of psychology may of course vary, due to the scholar we refer to, but especially during sixteenth and seventeenth centuries, psychology was very much a type of anthropology that included all the parts of the living human being (Vidal 2011). This included both the mind and the body, which implied that physiology could be regarded as belonging to psychology. Yet physiology was at the same time a part of nature, which for some led to the conclusion that psychology was a natural science. However, this did not contradict the fact that psychology was understood as anthropology, although that term was not applied as a name for a specific science. Psychology, in other words, was to be regarded as a broad science that included all aspects of living humans. To provide this aspect now, it is important to apply 'transdisplinarity' instead of 'interdisciplinarity' as the term.

On this basis, it is necessary to bring in different locations to see agency from a transdisciplinary perspective. Uher operates with three different properties that point in the direction of transdisciplinarity: the individual, the temporal and the physical. Those three locations reflect the areas: social science, humanities and natural science. In other words, those three academic main areas are very much united in psychology, and this is envisaged when an individual is regarded from a developmental perspective, and not least when the different aspects of the individual's genealogies are taken into account (Uher; Chap. 13).

This holistic perspective on psychology however does not disqualify psychology as being defined as an independent subject. Although some regarded psychology as a natural science, it was still a certain perspective on physiology that reflected a psychological angle. This perspective is first of all characterized by the subjective experiences of a lived life. This is crucial also when it comes to temporality. The psychological perspective on history is different from the historiographical in the sense that psychology focuses on experienced time instead of a common understanding of certain historical events. Psychology is also different from philosophy. The objective efforts of philosophy have never acknowledged the subjective perspective psychology focuses on. Yet in those cases where philosophy becomes subjective, psychology appears as more or less superfluous. However, despite the fact

that psychology can be regarded as an independent discipline, the transdisciplinary perspective is important in the effort of achieving explanatory power, especially when it comes to agency because agency is about acts performed in an environment covered by different disciplines. Thus, the axiom of transdisciplinarity:

> Every ecological perspective will by necessity transcend a certain discipline, although this does not imply a subversion of this discipline.

16.5 The Explanatory Power of Sign

The principle of transdisciplinarity is at the same time a principle of making connections. This stands often in contradiction to boundaries and distinctions. However, the reality may not necessarily be characterized as a fundamental contradiction, but rather as a world embedded with unavoidable dilemmas. This is very much what Charles Sanders Peirce took into account when he formulated the theorem: '[T]here is a character peculiar to every possible groups of objects' (Peirce 1878/1986b, p. 316). Although we may say that A=A is true, there will always exist a B, by which A=B also becomes true. This is the same as saying that the whole consists of related parts, which are detectable and can be specified and distinguished from each other, but still related.

Yet Peirce brings this theorem a step further by saying that the normal form of this theorem is, that '*every event must have a cause*' (Peirce 1878/1986b, p. 317). The cause, though, is not the blind efficient cause, but rather a relationship, which is mediated through the mind. This is the process of semiosis, which brings in a certain understanding of reality. In short, also 'a thought is in itself a reality' (Peirce 1872–73/1986, p. 78). This implies that the mind is not only a mediator of the reality but also a part of the reality itself. This is the basis for regarding semiosis as the way a human being handles the reality. And this is first of all by putting things together in terms of explanations by means of signs, because a sign 'is an object which stands for another to some mind' (Peirce 1872–73/1986, p. 66). Although the indexical sign is the most blatant example of the close relationship between sign and cause, all types of signs serve a causal function in the sense that the sign is the representation of an object and by this forms the object's cause.

On this basis, the semiosis is not only an infinite process but also a process in which the causal direction may vary and go both ways. The whole may be there as a result of the parts, but the parts may also be there because of the whole. And this is also how the social reality is to be understood, and not least how we can explain the Arab Spring in terms of aesthetical expressions (Awad and Wagoner; Chap. 14). If the production of sign 'redefines the situation, allowing the person to imaginatively project themselves beyond the here-and-now and thereby opening up for different possibilities of action' (Awad and Wagoner; Chap. 14, p. 229), we have a situation in which the future explains the present. This stands in contrast to the theory of planned behaviour because the meaning of the produced sign is not constituted by the intentional act alone, but rather by how the sign is perceived and received and by

this recreated in a communicative chain, which is continuously renewed and recreated. This is why Peirce stated, 'the existence of something in the present depends upon the future conditional occurrence of a certain event' (Peirce 1872–73/1986, p. 80). Thus, we cannot talk about the causes behind the Arabic revolution before we can take into account the conditions that actually tell us that this is a revolution, which the aesthetical expressions in terms of graffiti demonstratively seem to form (Awad and Wagoner; Chap. 14). Hence, the explanatory power of the sign appears to be very strong, and even maybe too strong because the criterion is so loose that it just requires a perceived and experienced relationship as the basis for talking about explanations. However, this is compensated by an unlimited need for demonstrations as the basis for reaching the level of having knowledge about something. Thus, we can formulate the axiom of sign:

> The explanatory power of the sign is so strong that it cannot be detached from meaning.

16.6 Conclusions

By this latter axiom, we are back where we started when we referred to Toulmin (1963). Our striving towards finding explanations is hard to separate from the process of meaning making. However, this is the psychological explanation for why we focus so much on explanations. The explanatory reasons we give for an event is on the other side not necessarily rooted in psychology. This is why philosophy of science is still an important and necessary contributor to the discussion about explanations. It provides the criteria for the validity of the argument. This is an important aspect in an evaluation of agency as an explanatory term in psychology. And what we have found in this respect is that (1) agency represents a contextualisation of intentionality, specifically in terms of coagency, (2) a full understanding of agency requires a transdisciplinary approach and (3) the semiotic approach does not make any clear distinction between explanations and meaning making, which implies that the scientific argument for agency is based on demonstrative specificity. Seen from a semiotic perspective however, 'demonstrative specificity' is also to be regarded as a sign.

References

Dilthey, W. (1996). *Hermeneutics and the study of history*. In R. A. Makkreel & F. Rodi (Eds.), (1985–2010) *Selected works/Wilhelm Dilthey* (Vol. 4). Princeton: Princeton University Press.

Elster, J. (1979). *Forklaring og dialektikk*. Oslo: Pax.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Gadamer, H.-G. (2004). *Truth and method*. London: Continuum.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory. Strategies for qualitative research*. New York: Aldine de Gruyter.

Heidegger, M. (1973). *Oikos og Techne. "Spørsmålet om teknikken" og andre essays. Idé og tanke*. Oslo: Tanum.

Klempe, S. H. (2014). Catalysis, functional explanations and functionalism in psychology. In K. R. Cabell & J. Valsiner (Eds.), *The catalyzing mind: Beyond models of causality* (Vol. 11). New York: Springer (Annals of theoretical psychology).

Max-Neef, M. A. (2005). Foundations of transdisciplinarity. *Ecological Economics*, *53*, 5–16. http://artsci.unsw.wikispaces.net/file/view/Max_Neef_Foundations_of_transdisciplinarity.pdf/340979004/Max_Neef_Foundations_of_transdisciplinarity.pdf.

Mengal, P. (2005). *La naissance de la psychologie*. Paris: L'Harmattan.

Peirce, C. S. (1872–73/1986). Toward a logic book. In Peirce, C. S. (Ed.), (1986). *Writings of Charles Sanders Peirce. A chronological edition* (Vol. 3, 1872–1878). Bloomington: Indiana University Press.

Peirce, C. S. (1878/1986). How to make our ideas clear. In Peirce, C. S. (Ed.), (1986). *Writings of Charles Sanders Peirce. A chronological edition* (Vol. 3, 1872–1878). Bloomington: Indiana University Press.

Peirce, C. S. (1878/1986b). Order of nature. In Peirce, C. S. (Ed.), (1986). *Writings of Charles Sanders Peirce. A chronological edition* (Vol. 3, 1872–1878). Bloomington: Indiana University Press.

Popper, K. R. (1980). *The logic of scientific discovery*. London: Hutchinson.

Toulmin, S. (1963). *The philosophy of science*. New York: Harper & Row.

Vidal, F. (2011). *The sciences of the soul. The early modern origins of psychology*. Chicago: University of Chicago Press.

Wundt, W. (1920). *Erlebtes und Erkanntes*. Stuttgart: Alfred Kröner.

Index

C. W. Gruber et al. (eds.), *Constraints of Agency,* Annals of Theoretical Psychology 12,
DOI 10.1007/978-3-319-10130-9