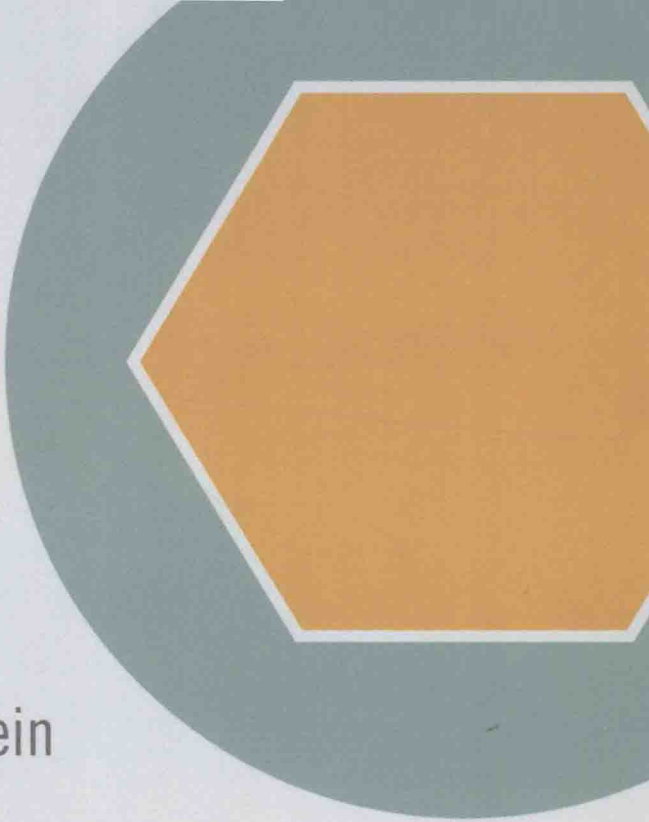

Stanley B. Klein

THE TWO SELVES

Their Metaphysical Commitments and Functional Independence

OXFORD

THE TWO SELVES

Their Metaphysical Commitments and Functional Independence

Stanley B. Klein

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Klein, Stanley B.
The two selves : their metaphysical commitments and functional
independence / Stanley B. Klein.
pages cm
Includes bibliographical references and index.
ISBN 978–0–19–934996–8
1. Self-consciousness (Awareness) 2. Knowledge, Theory of.
3. Metaphysics. I. Title.
BD450.K587 2013
126—dc23
2013020981

1 3 5 7 9 8 6 4 2
Printed in the United States of America
on acid-free paper

THE TWO SELVES

This book is dedicated to my mother, Julia Klein, and my father, Melvin Klein

PREFACE

Imagine a situation in which there are present three persons: an experimenter, an observer, and you. The experimenter tells you: "Lift your right index finger." You do so. After *N* seconds, the experimenter again gives the command, and again you obey. This procedure is repeated for a number of trials, over a 15-minute period, with *N* varying randomly from 2 to 30 seconds. Every time, the observer accurately predicts your behavior: you lift your finger when the command is given, not before it and always after it.

This sort of stimulus-response (S-R) connection is as a nice illustration of an instance ofnecessary and sufficient conditions for a specific behavioral act—lifting of a finger in a particular situation: Whenever S (the command to lift your finger) occurs, R (lifting your finger) occurs. In the absence of S, R does not occur. We have here a clear and simple demonstration of the validity of the utility of stimulus-response psychology.

Now, let's change the imagined scenario. You still are there, the same observer still is there, but the experimenter is not. The observer sees the same things that he saw before: Over a 15-minute

period, you lift your right index finger every now and then, with your responses randomly occurring at intervals between 2 and 30 seconds. However, this time the observer can no longer predict when you are going to lift your finger.

Here is the central question: "In this second scenario, *who* gives the command *to whom* to lift the finger?" One obvious way to capture what is going on is to suggest that the subject now is exercising his or her free will to engage in certain behavior. But this assertion simply describes the situation. It leaves little for further study. More important, to conflate—or just ignore the difference between—two clearly distinguishable aspects of the scenario, issuing a command and executing it, is counterproductive.

A preferable answer to the "who and to whom" question is that the situation can be conceptualized as the interaction between two distinctive components of the brain/mind. One of them "makes up" and issues the finger-lifting command, and the other "listens" and takes action. In this book, the two components of the mind/brain are conceptualized as two kinds of "self." "Self," as the reader probably knows, is the replacement of the earlier pre-scientific concept of the "homunculus." A voluminous literature on the self exists. As employed in contemporary psychology, the term *self* admits to a multitude of descriptions and meanings (e.g., Klein, 2004, 2010, 2012a). There are many ways of studying the self, but there is much confusion—as in the study of *consciousness*, a close family relation of self—although a reasonable degree of progress can be pointed to.

Of course, before speculating on the need to posit two aspects or kinds of selves (the "who" and "to whom") to account for the "finger-lifting scenario," determinist alternatives deserve consideration. Such explanations have the scientific "merit" of avoiding any need to postulate messy aspects of reality such as free will,

mind/body differences, or types of selves. A determinist might explain the second version of the scenario (i.e., finger lifting absent the experimenter's commands) by arguing that, while the finger movements were not causally related to the external commands, they were caused by something *other than* an interaction between your two selves. For example, Pierre-Simon Laplace's conception of determinism—according to which an infinitely intelligent demon who knows all the initial conditions and all of the physical laws relevant to your finger movements—suggests that your finger had *no choice* but to move when it did (of course, the phrase "no choice" might seem to reintroduce the question of free will, albeit via a dimly lit back door).

I think Laplace's demon argument—as well as its more recent philosophical variants—ultimately is unsatisfactory. Why? Because one of the premises of most determinist deductions—i.e., if we (a) know all the initial conditions, (b) know all the physically relevant laws, and (c) if we have if an intellect sufficiently sophisticated to subject this information to the correct analyses, then (d) we can predict with certainty the future state of any system—is, by Kurt Gödel's incompleteness theorem (e.g., Nagel & Newman, 2001; Rosser, 1936), false.

According to Gödel's theorem, one cannot supply proofs for all the laws that capture all the truths about *any* formal system from *within* that system. If this is so, then it is impossible to derive the internal consistency of a very large class of deductive systems (although Gödel's theorem was specifically targeted to arithmetic systems, it has been generalized to other internally consistent systems of axioms; for discussion as well as a critique, see Wang, 1996). In other words, for any given set of axioms, there are true mathematical statements that cannot be derived from the set itself: there will always be statements about the system that cannot

be proved within the system—hence the name "incompleteness theorem."

Mathematical statements that are assumed true, but cannot be proved within their system of origin, can, according to Gödel's logic, be proved in larger systems that can be shown to be valid forms of reasoning; they are simply undecidable in the more limited system. Thus, one always can seek a meta-formalization to capture all of the "truths" of any closed system. However, this incompleteness can be iterated infinitely. Accordingly, certain laws assumed true within a system of axioms cannot be proven within any finite time due to infinite regression. So, neither you, nor I, nor Laplace's demon can know with mathematical certainty *all* the laws. And if that is the case, then the deductive argument for determinism falls victim to the falsity of premise *b*.[1]

But that is not the only problem. Since precise specification of the fate of any determinist processes depends in a highly sensitive manner on exact knowledge of initial conditions of individual subatomic particles, and since, by the principle of quantum indeterminacy (see Chapter 3), knowledge of these conditions can never be obtained with sufficient precision, the equations of motion cannot be solved in an unambiguous manner. (Quantum mechanics makes precise predictions for the probabilities of the outcome of large aggregates of particles, and these probability distributions are deterministic; see, e.g., Thompson, 2008. However, predicting the fate of an individual particle remains indeterminable.) Thus, at the level of individual particles, premise *a* also cannot be realized, except by allowing for margins of error in prediction (which, by virtue of the dynamics of chaos theory, may be considerable).

While the issues raised do not necessarily entail that determinism is false (or that free will is true), neither do they provide much comfort to advocates of a fully deterministic reality. For my

purposes, the take-away message is that nothing in the sciences logically precludes the possibility that the "second scenario" of the finger-lifting exercise can be explained by postulating the activities (and causal potencies) of two selves. (Although questions concerning free will and determinism clearly are related to some of the core issues raised in this book, I will not deal with this large, contentious philosophical and psychological literature. We have more than enough to occupy us without opening that particular can of theoretical worms! The interested reader is referred to Balaguer, 2010; Kane, 2002; Libet, 1993; and Swinburne, 2011, for treatments of the role of free will in modern science, psychology, philosophy, and theology.)

The main message of this book is that the self is (a) real, (b) causally potent, and (c) consists of multiple aspects that have different parts to play in experience and behavior. A second, but important, message concerns the need to keep one's mind open to the *possibility* that reality, taken in its fullness, leaves room for aspects that do not admit to material instantiation. With regard to the self, I describe philosophical and psychological evidence in support of the idea that the self of everyday experience consists of two aspects—one material (the neuro-cognitive self) and one immaterial (the self of first-person subjectivity). I hope to show that each is an aspect of reality, that each constitutes a necessary condition for the human experience of "self" and that each has a causal role to play; however, I also argue that these two types of self have very different metaphysical commitments. Whether I succeed in making these points is something you will need to decide after reading what I have to say.

In Chapter 1, I make the case that the self is not, *contra* much psychological and philosophical doctrine, a "thing" to be studied. Rather, it is a multiplicity of aspects consisting of both

neural-cognitive (largely, but not exclusively, memory-based) instantiations as well as first-person subjectivity. I argue that the former aspect of self is material in nature and objectifiable, and thus amenable to scientific scrutiny. In contrast, the latter aspect of self is an immaterial subjectivity and thus not (easily) captured by the materialist dogma of modern science.

Chapter 2 describes more fully the neuro-cognitive aspect of self and presents evidence for the functional independence of its component systems (e.g., semantic self-facts, semantic trait self-knowledge, and episodic self-narratives). Much of this empiricism comes from my laboratory. For this, I apologize. I am one of the leading researchers on such matters (akin to being "king of the ant hill"—a rather small kingdom!), so the reader will have to suffer through an excess of "Klein" studies.

In Chapter 3, I describe the self of first-person experience and explore the *possibility* that this aspect of self might exist in non-material form. The position of modern science is that all of reality is ultimately material. However, as Meixner (2005) demonstrates in a carefully crafted analysis, materialism is a metaphysical position, not a scientific fact. The materialist stance is an example of what Rescher (1984) classifies as a scientific precommitment — that is, a *presumption* that helps determine the formative background of the questions we ask nature, rather than a *fact* we discover by virtue of the answers we receive (see also Hanson, 1958). Accordingly, a materialist stance does not have a greater claim on our credence than any other metaphysical position.

Many of the arguments I present in support of the *possibility* of immaterial aspects of reality draw on the principles of quantum indeterminancy and relativity theory. It is somewhat ironic, but curiously satisfying, that science itself provides some of the theoretical machinery and logical concepts needed to make discussion

of the possibility of a reality that supersedes the limitations of science feasible (as the reader may have noticed, this outcome has a decidedly Gödelian flavor!).

My arguments for the immateriality of the subjective aspect of self do not amount to a proof. Rather, they should be taken only as an appeal to broaden our appreciation of what might constitute reality. While I hope to convince you that here are things that the materialist stance of modern science cannot explain (e.g., Gendlin, 1962; Martin, 2008; Papa-Grimaldi, 1998), recognition of limitations is not the same as offering an alternative. But the recognition of those limits is a precondition to at least being open to alternatives (Nagel, 2012). One alternative, discussed in this book, is the possibility that there exists a conscious, immaterial aspect of reality—i.e., that consciousness, as exemplified by what I call the "ontological self" (see Chapter 1), is a central feature of nature rather than an epiphenomenon to be explained away via a materialist reductive analysis. Whether my arguments are sufficient to support this alternative approach to "what is real" is something you will have to judge for yourself. The best I can realistically hope for is that my arguments will convince you to leave open the metaphysical door, not that they will enable you to identify with certainty what passes through it.

In Chapter 4, I provide a brief—but, I believe, much needed—summary of the two aspects of self: the epistemological (i.e., the material) and ontological (i.e., the immaterial). My justification—if one allows me considerable leeway in the meaning and use of that word—for this unusual nomenclature for the two aspects of the self is provided in Chapter 1.

Chapter 5 considers arguments for the need for both psychological and material aspects of reality. I then present evidence from case studies (e.g., patients suffering anosognosia,

depersonalization, schizophrenic thought-insertion) showing that one can lose one's sense of personal ownership of one's mental states (e.g., "the thought/memory is in my head, but it is not mine!") while still maintaining a clear *sense* of one's material and immaterial self. This dissociation, I argue, hints that the "feeling of personal ownership" may be what unites these two metaphysically separable aspects of self.

In Chapter 6, I summarize the points made in previous chapters. I also appeal for a more inclusive approach to the empirical study of psychological reality—an approach that considers *all* aspects of experience as real, and attempts to understand those experiences using *all* the tools currently available. Finally, I conclude that the self of neural instantiation and the self of subjectivity are contingently related by personal ownership, and that this connection can be undone under certain conditions of pathology. This suggests, in turn, a functional independence of, and reality for, the two categories of self discussed in this book.

The possibility of an immaterial aspect of reality will, in all likelihood, be seen by readers as the most polarizing part of this book. The idea that an immaterial, conscious self might have a categorical irreducibility that is impossible to explain in terms of other categories of nature—such as mass, time, and space—has gained traction in recent years. Chalmers (1996), for instance, opined that a new ontology, in which consciousness is accorded the status of a fundamental aspect of reality, might be needed. But not everyone is comfortable granting the *immaterial* entry into a reality that is taken by modern science to be exclusively *material*. Materialist-minded folk (the majority of scientists and Western philosophers) will thus look away; those steeped in Eastern contemplative traditions will wonder what the fuss is about; and

theologians will find in the immaterial self a potential opening for discussion about the reality of the soul.

I agree with my theological friends that the possibility of an immaterial, consciousness self does have certain affinities with the concept of a "soul." However, I do not think the concept of immateriality explored in this book should be affiliated with any *particular* religious tradition, doctrine, or denomination. Immaterial self-consciousness, if it exists, may very well be capable of being related to a particular set of theological principles, but such a maneuver requires a rather Procrustean manipulation of the ideas presented herein.

ACKNOWLEDGMENTS

Many people have contributed both directly and indirectly to the ideas in this book. My deep gratitude goes to Moshe Lax, whose belief in my work combined with his financial support helped bring this book to fruition. Of great importance is Endel Tulving (who provided the "finger lifting" scenario that opens the Preface). I feel very fortunate to be able to consider Endel, in chronological order, a mentor, colleague, and friend. In terms of personal importance, the ordering would be reversed. Without Endel's support, encouragement, and feedback, I doubt seriously this project would have been undertaken.

I have benefited enormously from the insightful suggestions and gentle corrections from a truly gifted group of philosophers and psychologists. They generously gave of their time on what may have seemed to them a rather curious (and perhaps ill-conceived) project. Thanks are due to Robin Collins, David Funder, Byron Kaldis, Tim Lane, Marya Schechtman, Galen Strawson, and Richard Swinburne. I also want to express my gratitude to Carl Craver, Jonardon Ganeri, Shaun Nichols, and Dan Zahavi.

Although they were not directly involved in the writing of this book, our conversations have been invaluable for my understanding of the general issues involved.

I thank Joan Bossert and Miles Osgood, my wonderful editors at Oxford University Press. They put up with a lot, not the least of which was my continual request to substitute my "new and improved" version of the manuscript for the "new and improved version" of the previous day (despite all my "promises" to cease and desist from unending textual emendations)! While on the topic of "putting up", great appreciation goes to my wife, Judith Loftus, and my children, Nate and Eli, who had to endure a husband and father whose odd behaviors reached new levels as a result of my obsession with this project.

Finally, I want to stress that all of these individuals deserve credit for making my book better than it otherwise would have been. Any obscurities, weak arguments, and incoherencies that remain do so despite their best efforts to set me on the correct path.

THE TWO SELVES