

The
Journal of
Mind and Behavior

Vol. 37 No. 1 Winter 2016

ISSN 0271-0137

The *Journal of Mind and Behavior* (JMB) is dedicated to the interdisciplinary approach within psychology and related fields. Mind and behavior position, interact, and causally relate to each other in multi-directional ways; JMB urges the exploration of these interrelationships. The editors are particularly interested in scholarly work in the following areas: □ the psychology, philosophy, and sociology of experimentation and the scientific method □ the relationships among methodology, operationism, and theory construction □ the mind–body problem in the social sciences, psychiatry and the medical sciences, and the physical sciences □ philosophical impact of a mind–body epistemology upon psychology and its theories of consciousness □ critical examinations of the DSM–biopsychiatry–somatotherapy framework of thought and practice □ issues pertaining to the ethical study of cognition, self-awareness, and higher functions of consciousness in nonhuman animals □ phenomenological, teleological, existential, and introspective reports relevant to psychology, psychosocial methodology, and social philosophy □ historical perspectives on the course and nature of psychological science.

JMB is based upon the premise that all meaningful statements about human behavior rest ultimately upon observation — with no one scientific method possessing, a priori, greater credence than another. Emphasis upon experimental control should not preclude the experiment as a measure of behavior outside the scientific laboratory. The editors recognize the need to propagate ideas and speculations as well as the need to form empirical situations for testing them. However, we believe in a working reciprocity between theory and method (not a confounding), and in a unity among the sciences. Manuscripts should accentuate this interdisciplinary approach — either explicitly in their content, or implicitly within their point of view. (Note: we typically do not publish empirical research.)

JMB offers an immediate publication outlet on a quarterly basis. The Journal publishes one volume per year in the following sequence of issues: Winter, Spring, Summer, and Autumn. There are no page costs for accepted manuscripts. JMB is a refereed journal, and all decisions will be made by the assessing editors, associate editors, and chief editors. Commentaries and responses to individual articles and reviews are welcome. Authors subscribing at the time of manuscript submission are eligible for reduced subscription rates (see below).

All manuscripts should follow the style and preparation presented in the *Publication Manual of the American Psychological Association* (sixth edition, 2010). Particular attention should be paid to the citing of references, both in the text and on the reference page. [Note exceptions to APA style: JMB uses *no* ampersands or city/state abbreviations in referencing; the Journal uses three levels of headings: level 1, level 3, and level 4, see pp. 113, 114, 115 from the fifth (2001) edition APA Manual.] Authors requesting blind review must specify and prepare their manuscripts accordingly. Manuscripts may be sent to the Editor either by email to jmb@maine.edu or by post (one copy) to:

Raymond Chester Russ, Ph.D., Editor
The Journal of Mind and Behavior
Department of Psychology
University of Maine
5742 Little Hall
Orono, Maine 04469–5742
Tel. (207) 581-2057

Yearly subscription rates are \$32.00 for students or hardship; \$35.00 for past/present JMB authors or for those submitting manuscripts; \$46.00 for individuals; \$185.00 for institutions. Air mail rates upon request. All back issues are available and abstracts are located at www.umaine.edu/jmb. For detailed information contact our Circulation Department at The Institute of Mind and Behavior, P.O. Box 522, Village Station, New York City, New York 10014. Tel: (212) 595-4853.

The Journal of Mind and Behavior

Editorial Board

Editor:

Raymond Chester Russ, Ph.D.
Department of Psychology
University of Maine

Associate Editors:

Charles I. Abramson, Ph.D.
Department of Psychology
Oklahoma State University

David Cohen, Ph.D., MSW
Department of Social Welfare
UCLA Luskin School
of Public Affairs

David H. Jacobs, Ph.D.
Private Practice
San Diego, California

Thomas Natsoulas, Ph.D.
Department of Psychology
University of California, Davis

Richard D. Schenkman, M.D.
Private Practice
Bakersfield, California

Laurence Smith, Ph.D.
Department of Psychology
University of Maine

Book Review Editor:

Steven E. Connelly, Ph.D.
Department of English
Indiana State University

Editorial Board Liaison and Review Editor:

Leslie Marsh, Ph.D.
International Academy of Pathology
Faculty of Medicine
University of British Columbia
Vancouver, Canada

Creative Director:

Kate McPherson
Yarmouth, Maine

Liaison for Medical Affairs:

Elliot M. Frohman, Ph.D., M.D.
Department of Neurology and
Ophthalmology
Southwestern Medical
Center at Dallas
University of Texas

Assessing Editors

Kenneth Aizawa, Ph.D.
Department of Philosophy
Centenary College of Louisiana
Shreveport, Louisiana

Kristin Andrews, Ph.D.
Department of Philosophy
York University
Toronto

John Antrobus, Ph.D.
Department of Psychology
The City College of New York

James Bailey, Ph.D.
Graduate School of Management
Rutgers University
Newark, New Jersey

Ken Barney, M.D.
Cambridge, Massachusetts

Amanda J. Barnier, Ph.D.
ARC Centre of Excellence
in Cognition and its Disorders
Macquarie University
Sydney, Australia

Clare Batty, Ph.D.
Department of Philosophy
University of Kentucky
Lexington, Kentucky

Hagit Benbaji, Ph.D.
Philosophy Department
Ben-Gurion University of the Negev

Mark Blagrove, Ph.D.
Department of Psychology
University of Wales Swansea

Richard Booth, Ph.D.
Department of Social and
Behavioral Studies
Black Hawk College

Robert F. Bornstein, Ph.D.
Department of Psychology
Gettysburg College

Gregg D. Caruso, Ph.D.
Department of Philosophy
Corning Community College, SUNY
Corning, New York

Anthony Chemero, Ph.D.
Departments of Philosophy
and Psychology
University of Cincinnati

Paul D. Cherulnik, Ph.D.
Leeds, Massachusetts

Phyllis Chesler, Ph.D.
Department of Psychology
College of Staten Island, CUNY

Mazviita Chirimuuta, Ph.D.
Department of History and
Philosophy of Science
University of Pittsburgh

Juan J. Colomina-Almiñana, Ph.D.
Director, Program of Language
and Cognition
Department of Mexican
American and Latina/o Studies
University of Texas, Austin

Dr. Jean-Pierre Courtial
Laboratoire de Psychologie
Université de Nantes

Katja Crone, Ph.D.
Department of Philosophy and
Political Science
Technische Universität Dortmund

Mark Crooks
Department of Psychology
Michigan State University

Paul F. Cunningham, Ph.D.
Dean, Liberal Arts and Sciences
Rivier University
Nashua, New Hampshire

Edward Dale
Stockton Hall Psychiatry Hospital
Stockton-on-the-Forest, England

Helen De Cruz, Ph.D.
Department of Philosophy
VU University Amsterdam

Helena De Preester, Ph.D.
Department of Philosophy and
Moral Science
Ghent University
Ghent, Belgium

Florence L. Denmark, Ph.D.
Psychology Department
Pace University

Susannah Kate Devitt, Ph.D.
Information Systems School
Science and Engineering Faculty
Queensland University of Technology

James Dietch, M.D.
California College of Medicine
University of California, Irvine

Michael C. Dillbeck, Ph.D.
Department of Psychology
Maharishi University
of Management

Larry Dossey, M.D.
Private Practice
Santa Fe, New Mexico

Paula Droegge, Ph.D.
Department of Philosophy
Pennsylvania State University

Włodzisław Duch, D.Sc., Ph.D.
Department of Informatics
Nicolaus Copernicus University
Toruń, Poland

Monica L. Duchnowski, Ph.D.
West New York, New Jersey

Matthew G.N. Dunlap
Maine Secretary of State
Augusta, Maine

Arthur Efron, Ph.D.
Department of English
SUNY at Buffalo

Robert Epstein, Ph.D. Cambridge Center for Behavioral Studies	Marcelino Guillén, LCSW Aging Well Life Care Bronx, New York	Rebecca Kukla, Ph.D. Department of Philosophy University of South Florida
Emily Esch, Ph.D. Department of Philosophy College of Saint Benedict and Saint John's University Collegeville, Minnesota	William L. Hathaway, Ph.D., Director Doctoral Program in Clinical Psychology Regent University	James T. Lamiell, Ph.D. Department of Psychology Georgetown University
Simon Evnine, Ph.D. Department of Philosophy University of Miami Coral Gables	Jeffrey Hershfield, Ph.D. Department of Philosophy Wichita State University	Jane Suilin Lavelle, Ph.D. School of Philosophy, Psychology, and Language Sciences University of Edinburgh
Seth Farber, Ph.D. Network Against Coercive Psychiatry New York City	Robert R. Hoffman, Ph.D. Florida Institute for Human and Machine Cognition Pensacola, Florida	Wendy Lee, Ph.D. Department of Philosophy Bloomsburg University
James Fastook, Ph.D. Department of Computer Sciences University of Maine	Manfred J. Holler, Ph.D. Institute of SocioEconomics Munich	Dorotheé Legrand Centre National de la Recherche Scientifique École Normale Supérieure Paris
Carrie Figdor, Ph.D. Department of Philosophy and Interdisciplinary Graduate Program in Neuroscience University of Iowa Iowa City, Iowa	Daniel D. Hutto, DPhil School of Humanities and Social Inquiry University of Wollongong Australia	Jonathan Leo, Ph.D. Department of Anatomy Lincoln Memorial University
Owen Flanagan, Ph.D. Department of Philosophy Duke University	Gabrielle Benette Jackson, Ph.D. Department of Philosophy Stony Brook University	Hemdat Lerman, Ph.D. Department of Philosophy University of Warwick Coventry, United Kingdom
Tom Froese, D.Phil. Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas Universidad Nacional Autónoma de México, Mexico City	Marta Jorba, Ph.D. Department of Philosophy University of Girona Spain	JeeLoo Liu, Ph.D. Department of Philosophy California State University Fullerton
Liane Gabora, Ph.D. Department of Psychology University of British Columbia	J. Scott Jordan, Ph.D. Department of Psychology Illinois State University	Michelle Maiese, Ph.D. Department of Philosophy Emmanuel College Boston
Nivedita Gangopadhyay, Ph.D. Department of Philosophy University of Bergen Bergen, Norway	Jay Joseph, Psy.D. Private Practice Berkeley, California	Pete Mandik, Ph.D. Department of Philosophy William Paterson University
Kenneth J. Gergen, Ph.D. Department of Psychology Swarthmore College	Justine Kingsbury, Ph.D. Philosophy Programme University of Waikato Hamilton, New Zealand	Frank McAndrew, Ph.D. Department of Psychology Knox College
Grant R. Gillett, D. Phil. (Oxon.), M.D. Department of Philosophy University of Otago, New Zealand	Andrzej Kokoszka, M.D., Ph.D. Department of Psychiatry Jagiellonian University Kraków, Poland	Michael Montagne, Ph.D. Massachusetts College of Pharmacy Boston
Brian Glenney, Ph.D. Department of Philosophy Gordon College Wenham, Massachusetts	Paul Krassner, Editor The Realist Venice, California	Alain Morin, Ph.D. Department of Behavioral Sciences Mount Royal College Calgary
Aaron D. Gresson, III, Ph.D. Center for the Study of Equity in Education Pennsylvania State University	Stanley Krippner, Ph.D. Saybrook Graduate School Oakland	Ezequiel Morsella, Ph.D. Department of Psychology San Francisco State University
Ronald P. Gruber, M.D. Clinical Associate Professor University of California (SF) Stanford University Medical Center	Gerhard M. Kroiss, Ph.D. International Institute for Critical Thinking Greensboro, North Carolina	Jennifer Mundale, Ph.D. Department of Philosophy University of Central Florida Orlando, Florida
	Joel Krueger, Ph.D. Department of Sociology, Philosophy, and Anthropology University of Exeter Exeter, United Kingdom	Paul G. Muscari, Ph.D. Department of Philosophy State University College of New York at Glens Falls

Raymond A. Noack
CMS Research
Seattle, Washington

Dr. Christian Onof
Department of Philosophy
Birkbeck College
London

Nico Orlandi, Ph.D.
Department of Philosophy
University of California at Santa Cruz

Kenneth R. Pelletier, M.D., Ph.D.
School of Medicine
Stanford University

Trevor Persons, Herpetologist
USGS Colorado Plateau
Research Station
Northern Arizona University
Flagstaff, Arizona

Gerard A. Postiglione, Ph.D.
School of Education
University of Hong Kong

Isaac Prilleltensky, Ph.D.
Department of Human and
Organizational Development
Vanderbilt University

Rachel Naomi Remen, M.D.
Saybrook Graduate School
Oakland

Louise Richardson, Ph.D.
Department of Philosophy
University of York

Rochelle P. Ripple, Ed.D.
Department of Education
Columbus State University
Columbus, Georgia

Sarah Robins, Ph.D.
Department of Philosophy
University of Kansas
Lawrence, Kansas

Komarine Romdenh-Romluc, Ph.D.
Department of Philosophy
University of Sheffield
United Kingdom

Steven Rosen, Ph.D.
Department of Psychology
College of Staten Island, CUNY

Ralph L. Rosnow, Ph.D.
Department of Psychology
Temple University

Jeffrey Rubin, Ph.D.
Psychology Department
Corning-Painted Post Area
School District

Robert D. Rupert, Ph.D.
Department of Philosophy
University of Colorado
Boulder, Colorado

J. Michael Russell, Ph.D.
Department of Philosophy
California State University
Fullerton

Joshua Rust, Ph.D.
Department of Philosophy
Stetson University
DeLand, Florida

Henry D. Schlinger, Ph.D.
Department of Psychology
California State University
Los Angeles

Maureen Sei, Ph.D.
Department of Philosophy
Erasmus University and
Leiden University
The Netherlands

Bernard S. Siegel, M.D.
Surgical Associates of
New Haven
New Haven, Connecticut

Laurence Simon, Ph.D.
Kingsborough Community
College
Brooklyn, New York

Janusz Slawinski, Ph.D.
Institute of Physics
Pedagogical University
Kraków, Poland

Brent D. Slife, Ph.D.
Department of Psychology
Brigham Young University

Tonu Soidla, Ph.D., D.Sc.
Institute of Cytology
St. Petersburg, Russia

Steve Soldinger, M.D.
Neuropsychiatric Institute
University of California
Los Angeles

David Spurrett, Ph.D.
Discipline of Philosophy
University of KwaZulu-Natal
South Africa

Peter Stastny, M.D.
Private Practice
New York City

Lincoln Stoller, Ph.D.
Mind, Strength, Balance Ltd.
Shokan, New York

Liz Stillwaggon Swan, Ph.D.
Philosophy Department
Mercyhurst University
Erie, Pennsylvania

Şerife Tekin, Ph.D.
Department of Philosophy
and Religious Studies
Daemen College
Amherst, New York

Nigel J.T. Thomas, Ph.D.
Division of Humanities and
Social Sciences
California Institute of Technology

Deborah Perron Tollefsen, Ph.D.
Department of Philosophy
University of Memphis

Warren W. Tryon, Ph.D., ABPP
Department of Psychology
Fordham University

Larry Vandervert, Ph.D.
American Nonlinear Systems
Spokane, Washington

Wayne Viney, Ph.D.
Department of Psychology
Colorado State University

Glenn D. Walters, Ph.D.
Psychology Services
Federal Correctional Institution
Schuylkill, Pennsylvania

Duff Waring, L.L.B.
Department of Philosophy
York University

Daniel A. Weiskopf, Ph.D.
Department of Philosophy
Georgia State University
Atlanta, Georgia

Richard N. Williams, Ph.D.
Department of Psychology
Brigham Young University

Jennifer M. Windt, Ph.D.
Philosophy Department
Monash University
Melbourne, Australia

Fred Alan Wolf, Ph.D.
Consulting Physicist
La Conner, Washington

Cory Wright, Ph.D.
Department of Philosophy
California State University
Long Beach

Kritika Yegnashankaran, Ph.D.
Program of Philosophy
Bard College
New York

Helen Yetter-Chappell, Ph.D.
Department of Philosophy
University of York
United Kingdom

Robert C. Ziller, Ph.D.
Department of Psychology
University of Florida

The
Journal of
Mind and Behavior

Vol. 37 No. 1

Winter 2016

Library of Congress Cataloging in Publication Data

The Journal of Mind and Behavior. – Vol. 1, no. 1 (spring 1980)–
– [New York, N.Y.: Journal of Mind and Behavior, Inc.] c1980–

1. Psychology–Periodicals. 2. Social psychology–Periodicals. 3. Philosophy–Periodicals.I.
Institute of Mind and Behavior
BF1.J6575 150'5 82-642121
ISSN 0271-0137 AACR 2 MARC-S

Copyright and Permissions: ©2016 The Institute of Mind and Behavior, Inc., P.O. Box 522, Village Station, New York City, New York 10014. All rights reserved. Written permission must be obtained from The Institute of Mind and Behavior for copying or reprinting text of more than 1,000 words. Permissions are normally granted contingent upon similar permission from the author. Printed in the United States of America.

The Journal of Mind and Behavior

Winter 2016

Vol. 37 No. 1

CONTENTS

Which Identification is Disturbed in Misidentification Syndromes? A Structural Analysis of Fregoli and Capgras Syndromes	1
Stéphane Thibierge and Catherine Morin	
Does Functionalism Offer an Adequate Account of Cognitive Psychology?	15
James M. Stedman, Thomas L. Spalding, and Christina L. Gagné	
Consciousness is Not a Physically Provable Property	31
Catherine M. Reason	
On the Appearance and Reality of Mind	47
Demian Whiting	

Which Identification is Disturbed in Misidentification Syndromes? A Structural Analysis of Fregoli and Capgras Syndromes

Stéphane Thibierge

Université Paris-Diderot

Catherine Morin

Université Pierre et Marie Curie

Based on a structural reading of the first observations of Fregoli syndrome by Courbon and Fail in 1927, of Capgras syndrome by Capgras and Reboul-Lachaux in 1923, as well as two present-day cases, we show that the essential feature of Fregoli syndrome is the disjunction between *recognition* and *identification*, two terms that are far from being synonymous. Fregoli syndrome is not just of historical interest to today's clinicians: it also allows us to separate out certain fundamental elements of what is ordinarily called recognition, elements that appear only in more indirect ways and latent forms in neurosis and in everyday psychopathology. The analysis of this syndrome therefore gives us access to the various elements of the matrix function for representation that Lacan described under the term *specular knowledge*.

Keywords: Fregoli syndrome, identification, recognition

In current psychiatric literature, Fregoli and Capgras syndromes are seen as rare psychotic syndromes. Together with intermetamorphosis and the syndrome of subjective doubles, they are part of the group of Delusional Misidentification Syndromes (DMS) and are considered a disturbance in *recognizing* or *identifying* people, the two terms being employed as synonyms. We question the implicit assumption that has allowed, on the one hand, to characterize Capgras and Fregoli syndromes as recognition disorders and, on the other hand, to classify

We are indebted to Dr. Marcel Czermak for having first drawn our attention, through his original teaching, to the clinical and theoretical interest of the Fregoli syndrome. Our thanks also go to Dr. Henry Frignet for his valuable bibliographical suggestions and to Kristina Valentinova who reread the English. All correspondence regarding this article should be addressed to Stéphane Thibierge, 11, rue Nicolas Charlet, 75015 Paris, France. Email: stephane.thibierge@wanadoo.fr

them as misidentification syndromes. From a psychoanalytic point of view, the assumption that identification and recognition are identical processes is in fact far from being self-evident. In claiming that the very same persecutor is disguised as many different people, or that persecutors have taken on the appearance of the patient's relatives, Fregoli and Capgras sufferers clearly show us that recognition and identification are two separate processes. Although the same analysis could easily be applied to intermetamorphosis and the delusion of subjective doubles, the present paper will focus on the Fregoli and Capgras delusions. The question of the significance of delusional misidentifications has been addressed by Cutting (1991) and Margariti and Kontaxakis (2006), who proposed that the common feature of DMSs (whether or not they involve recognition of people) was a disorder of identity or uniqueness. In the present paper, we show that the perception of the "uniqueness" of persons or objects depends on complex relations between recognition and identification.

Fregoli and Capgras syndromes were described by French psychiatrists in the 1920s under the generic term of false-recognition illusions of the insane (*illusions de fausse reconnaissance des aliénés*).¹ This terminology was used to differentiate the disorders from, on the one hand, ordinary false recognition — i.e., mistaking a person for somebody else due to an error or absent-mindedness — and, on the other hand, neurological deficits, for example, those affecting memory. While Capgras considered these disorders as manifestations of what he called "systematic misrecognition" (*méconnaissances systématiques*), the term "false recognition" (*fausses reconnaissances*) has mainly prevailed. Three syndromes became nosographic references: the *syndrome d'illusion des sosies* or syndrome of subjective doubles (Capgras and Reboul-Lachaux, 1923), the *syndrome d'illusion de Fregoli* or Fregoli syndrome (Courbon and Fail, 1927) and finally, the *syndrome d'intermétamorphose* or intermetamorphosis syndrome (Courbon and Tusques, 1932). The departure point for the work of these psychiatrists was Capgras's description of a symptom he found in one of his patients, Mrs. M., who suffered from persecution megalomania.

The patient maintained that her children had been stolen, hidden in the underground of Paris, and that her husband and her daughter had been replaced by multiple sosies. These sosies looked like her relatives but there were small differences. The case observation yielded other important elements. The patient gave herself a variety of proper names; she claimed that she was called Madame de Rio-Branco and was a descendant of numerous prestigious figures from a range

¹This is not to say that the phenomena identified by these syndromes had not been described before. To our knowledge, the oldest known description can be found in Leuret's *Fragments Psychologiques sur la Folie* (1834, pp. 115–118). A patient addresses Leuret and another doctor in the following terms: "You transform yourself," she would say — and Leuret would ask her against which one of them both these reproaches were directed. "It is you," she would then reply, "it makes only one, it is the same person."

of historical eras. With such a glorious ancestry, Mrs. M. also had an enormous fortune and throughout the centuries had given birth to an extraordinary number of children, all of whom had been stolen from her, replaced by doubles and hidden in mysterious places. In case that she herself might be replaced by a double, the patient wrote a description supposed to allow people to recognize her. This description involved a few anthropometric indices, but the patient mostly described her clothes with various details, her habits (for example, that she was normally accompanied by her daughter), and gave her address. To characterize Mrs. M.'s most striking symptom — her belief that her close relations had been replaced by sosies — Capgras coined a new term, *agnosie d'identification* (identification agnosia); the term was not used outside the context in which it was introduced.

In 1927, Paul Courbon and Gabriel Fail identified the Fregoli syndrome in a “schizophrenic” patient. They borrowed the name “Fregoli” from the words used by the patient, who maintained that her main persecutor, the actress Robine, was able to embody multiple different characters, just like the famous Italian actor Fregoli. The patient thus saw Robine in the people she met; to her they were all Robine in disguise. However, the patient never said that these disguised figures had identical faces. Instead she insisted that although their appearances differed, they were always the same person, “a single being” (Courbon and Fail, 1927, p. 123), who was responsible for a variety of phenomena imposed on her against her will — powers of magnetism, passionate outbursts, obscene commands, etc. These commands included an obligation to masturbate. The patient believed that while destroying her own body, these imposed acts simultaneously created an attractive dark line around Robine's eyes. The patient's right index finger, through which Robine increased her own beauty, was therefore worth several millions francs. The patient eventually tried to attack one of the figures she identified as Robine.

Although, today, Fregoli, Capgras and other delusional misidentification syndromes are considered to be rare, as Mojtabai (1998) has pointed out, their frequency may be underestimated. Some important aspects of Fregoli syndrome can be observed in cases that are not classified as either misidentification syndromes or monothematic delusions (Mojtabai, 1994). Indeed, some key features of the DMS syndromes are part of any psychosis. For example, in *Memoirs of My Nervous Illness* (Schreber, 2000, pp. 99–114), a self-report of a paranoiac delirium, we find some vivid examples of what appears very similar to the Fregoli and intermetamorphosis syndromes. As mentioned above (see footnote 1), Leuret had described a Fregoli patient as early as 1834. Although very few current observations give us access to the patient's speech, we have found two cases reported by French psychoanalysts (Thibierge and Morin, 2010). The first one comes from Porge (1986); the second is our own case (Thibierge, 2011).²

²For cases of other misidentification syndromes see Breen, Caine, Colheart, Hendy, and Roberts (2000) and Caine (2009).

Porge described the case of a female patient, who had felt at a glance that she loved a man, “Peter.” She met him once thereafter, but was not sure that it was really him; she thought it could have been someone else. After the second meeting, she kept thinking about Peter and came to believe he wanted her to reach him. Indeed, she “saw” him in the various other men she met. Each time, she was certain that this was Peter, but Peter as a transformed person. She was sure because she felt attracted to these different men. Moreover, she felt that Peter had stolen one of her lips: he controlled her upper lip when she was speaking (the lower lip remained hers), so that she was obliged to speak “with a small voice,” which was also Peter’s voice. In this case, it is always the one and the same person, Peter, who takes on the appearance of other men (the men she loves), who commands (attracts) her and steals a part of her body (her upper lip). As in Courbon and Fail’s *princeps* case, the patient identifies the same “magnetic” personality behind the appearance of several different people and this personality commands power over particular parts of her body.

Thibierge (2011) reported the case of a female patient with paranoid personality disorder, whose persecutor did not exactly take on the form of other people, but managed to alter the patient’s own appearance or to make other persons appear. The patient described how her life had been made difficult by a long-standing conflict with a particular nun, who was once one of her childhood teachers and had been persecuting her ever since. The nun would sometimes make different people enter the patient’s house without being seen by anybody but her. She caused the patient to be mistaken for someone else in the small town where she lived. The patient could hear the nun’s voice giving orders to her husband or members of her family, or even speaking from various parts of the patient’s body (hand, eye, foot). The voice also talked to people on television, making them turn their faces towards the patient. The nun thus changed her appearances and names; she impersonated various other persons, sometimes taking the place of the patient herself. She used her voice or her gaze to guide and control what the patient called *les humaines* — the human appearances of persons. In this case, as in the one reported by Porge (1986) and in the *princeps* case, the same being causes various changes in human appearance, commands those appearances, and partially controls and takes possession of the patient’s body.

The Fregoli and Capgras Syndromes: A Lacanian Perspective

Fregoli patients identify the people they meet as the persecutor in disguise; they receive from the persecutor various “influxes” or fluids, as well as sensorial phenomena that are imposed on them against their will. They recognize the individuals they meet as having different appearances, but they identify them as always being the same personality with the same name (Robine, Peter, the nun, etc.), that is, patients tend to identify always the same “figure” beneath the range of

people that they meet. As Courbon and Fail write in their first observation of the syndrome, Fregoli is “a single being” (1927, p. 123). The Fregoli patients maintain the following belief: the image may change — they know that the appearance is not the same — but in fact, it is so-and-so, it is really him or her, always the same, pursuing the patient. He is not *alike*, but *the same*. Hence we could say that, according to the patient’s words, *the other is always the same*. This phenomenon can be considered as a logical variation on the subjective doubles in the Capgras syndrome, where the subject recognizes someone, but cannot conclude as to his or her identity: in reality, it is not exactly him or her, it is a double. We could say that in this case *the same is always an other*. Fregoli and Capgras syndromes obviously have to do with the recognition and identification of persons. In Lacanian terms, one could say that these syndromes disturb the processes involved in *recognizing* somebody’s *image* and giving it a *name*. Although recognition is not currently considered as part of the field of psychoanalytical research but rather of cognitive science, psychoanalysis offers us specific and valuable tools to understand the relations between body image and nomination. It is worth recalling that Freud (1919) used the term “uncanny” to describe a disturbance of recognition. Moreover, in his *Project for a Scientific Psychology* (1895), Freud claimed that the normal recognition process should actually be thought of as an attempt to find again an object (*das Ding*) he considered to be fundamentally lost. Subsequently, Lacan made an indirect but invaluable contribution to our understanding of the nature of recognition, in his “Remarks on Daniel Lagache’s Presentation” (1966/2002d) and his seminar on “Anxiety” (2004). Recognition refers to everything that, without receiving special attention, presents itself to us as reality. This was concisely formulated by Lacan in his dialogue with Henri Ey (Lacan, 1966/2002c, p. 130): “For there is no antinomy whatsoever between the objects I perceive and my body, whose perception is constituted by a quite natural tuning with those objects.” When this “quite natural tuning” fails, we are faced with an order of facts we cannot recognize as ordinary reality. This manifests itself through feelings ranging from a fleeting discomfort barely tinged with anxiety — such as when we no longer recognize a familiar word while reading — to a complete collapse of reality.

While normal recognition makes reality self-evident to the point of preventing its analysis, psychotic disorders, which regularly involve a variety of syndromes isolated under the term *feelings of strangeness*, give us a unique opportunity to study disturbances of recognition. Fregoli and Capgras syndromes in particular display a decomposition of the basic elements involved in the process of recognition. This decomposition makes it possible for the two different dimensions of recognition to appear as separate from one another: the *name* on the one hand, the *image* on the other. In the words of these patients, the name designates something that the image fails to cover or represent, i.e., something that cannot be recognized. Indeed the main goal of any recognition process is to

grasp an image. And yet, what is at stake in Fregoli and Capgras syndromes is clearly something different.³

The image that Fregoli patients are confronted with has lost its consistency and identity of form. This follows modalities that may range from a unifying conjunction with the persecutor to a disjunction from the persecutor via a fragmentation of the body. In these syndromes, we could say that the patients do not actually deal with images, but rather with *something else*, something that has taken on a persecutory tone. Capgras and the subsequent psychiatrists interested in the syndrome seem to have found it difficult to characterize this something else using concepts related to recognition, so much so that they used a variety of terms — *méconnaissance* (misrecognition), *agnosie d'identification* (identification agnosia), *identification délirante* (delusional identification). Lacanian theory proves very useful in helping us characterize this *something else* and its relation to both name and body image. For the time being, we will denote this element, this something else that the patient designates as always the same, simply “*x*.” As an hypothesis, we suggest that this *x* is the cause of the disjunction between the name and the image.

Courbon and Fail’s observation makes it possible to pinpoint another element clearly revealed by Fregoli syndrome, namely that these phenomena cannot be understood solely as a defect in the field of recognition, and especially not bodily recognition. On the contrary, recognition, and especially recognition of the body image, is in this case fragmented, broken up into its component parts, for the benefit of something that the patient gives a name to, and hence positively identifies. This is obvious in the *princeps* observation: not only does the patient acknowledge that there is a variety of images — these are the others she comes across, meets in the street — but also these different images draw her attention by all sorts of imposed phenomena, — magnetic powers, outbursts, obscene commands, etc. — all of which she relates to the actress Robine. In other words, she identifies the same *x* in each one of them and names it by saying: “It is Robine.” This *x* identification is associated with a fragmentation of the representation of her own body. The breaking up of the patient’s body image is attested to by her speaking of her finger as a separate object, a merchandise: she says that her right index finger, which Robine shamelessly uses to increase her own beauty, is worth several millions francs. Robine’s beauty is thus linked to the destruction and fragmentation of the patient’s own body. In other words, there is her own fragmented image on the one side and what she calls *Robine* on the other side. Likewise, in the case described by Porge (1986), the patient’s lower lip is owned by Peter. The patient examined by Thibierge (2011) has a part of her limbs controlled by her persecutor.

³What we assume here does not involve only recognition of persons, but any recognition process. It is worth noting that in intermetamorphosis syndrome, disorders in recognizing people may be associated with disorders in recognizing objects or places.

Based on this, we can formulate the question as follows: What is the x that the patient identifies behind the figures she sees? Taking into account what the patient says, x appears to be an object that is:

1. autonomous — following on its own causality,
2. xenopathic — imposing various sensorial phenomena on the patient,
3. at the origin of the disintegration of the body image, and
4. one (always the same).

In place of the image and instead of it, the patient always identifies *the same*. But what is this sameness? Does it mean the same person? The same name? The same image? The same thing? Indeed we have difficulties trying to answer these questions — our ordinary logic does not suffice to characterize this sameness. This x , always the same but disguised in various forms, cannot be defined in cognitive terms. However, the characteristics which we have briefly summarized above fit in precisely with what Lacan referred to as the “object a .” A basic premise here is that the object a is not something that can be designated in reality. It is what Freud (1915) emphasized as the repressed drive impulses or instinctual representatives: having been repressed, these impulses or representatives cause the subject’s desire, without any possibility for them to be directly and positively identified. The object of these repressed impulses is what one yearns for and aims at through one’s demands, actions, dreams, symptoms, Freudian slips, etc. But it can never be grasped as such. In psychoanalytical theory, it is defined as the “lost object,” insofar as it is fundamentally linked to the repression involved in the subject’s entrance into language. Lacan designates it by a simple letter, a , in order to emphasize that it results from the loss inherent to language — the loss of any direct relation to what desire aims at. This is also why it cannot be objectively defined: it has no more objective meaning or form than a letter of the alphabet.

Lacan’s analysis based on his concept of the mirror stage (Lacan, 1966/2002d) enabled him to show very clearly that the image of one’s body, or *specular image*, can only take on a recognizable form and consistency on the condition of representing the loss, the absence from reality, of what he later called the object a . It is beyond the limits of this paper to discuss this concept in detail (for a more thorough discussion of the body image see Morin and Thibierge, 2006). In order to recognize one’s own image or that of another person, the subject must first be able to grasp it as a symbol, i.e., as indicating the loss or absence of something. Indeed, every symbol implies the possible absence of the symbolized object. Schematically speaking, we could assume that we can only recognize our own image on the condition that the object a has been repressed.

Lacan (1966/2002d) had this relation between image and object in mind when he proposed the formula of the body image: $i(a)$. The formula designates the image i as deriving its consistency from an object, the object a , whose absence is enveloped by the image. In terms of Lacan’s theoretical development, we could say that the formula encapsulates the entire process leading him from

Aimée, the patient to whom he devoted his doctoral thesis and who taught him to identify a constant element present throughout the series of her persecutor's images, to the "optical schema," where the real object derives its efficacy from being invisible (Lacan, 1966/2002d), and including the mirror stage (Lacan, 1966/2002a, 1966/2002b), where the illusory yet thoroughly captivating aspect of body image is brought to light. It may seem paradoxical and even provocative to assume that the image envelops the absence of the object. However, we all know that the image of our body may at times become concentrated in a kind of anxious attention, one that can never be satisfied, indeed as if there was always something missing. It is because our body image symbolizes and represents a lack that it is often experienced as unsatisfactory, incomplete, or downright strange. The Fregoli syndrome reveals this structure of the body image in a uniquely pure form by breaking it up: on one side, we have the unravelled, dislocated and unrecognizable image; on the other side, the object, which, in this case, is neither repressed nor lacking, but instead identified by the patient.

We can thus understand the formula $i(a)$ as precisely referring to the knotting together of the elements that delusional false recognitions reveal in a clearly isolated state. Though it is not part of the formula $i(a)$, the proper name is also altered in Fregoli and Capgras syndromes. The identified object, which repeatedly intrudes on the patient's attention, is given a unique, unequivocal name (Robine, Peter, the nun in Fregoli syndrome, the sosies in Capgras syndrome).⁴ Although it sounds like the persecutor's name, this unique name designates all the avatars of the persecutor, thus functioning as a common rather than proper name. Its efficacy extends beyond the usual function of the name; it names something that has the property of coming back to the subject under the guise of a real and unequivocal identity. This x designates precisely what the above formula designates as a , i.e., the object. However, while in neurosis this object is in principle never identified by the subject, here it is identified and even constitutes the mainspring of the systematization of the delusion. Not only is the name reduced to the status of a common name, but it is also further reduced to the object. In this transformation, the name loses its effective power, its identification capacity. In language, a name can in fact only identify through its difference from other names. It does not actually have a direct connection to what it names — except possibly in psychosis, as is the case here. From a clinical point of view, it is really one of the most distinctive features of this syndrome that the name actually connects itself to the object while identifying it. The Fregoli syndrome

⁴This loss of differentiation might be related to the basic characteristic of psychosis that Czermak (1986) brought to light in his pioneering work on Cotard's syndrome, under the term of "de-specification" of the body's orifices. In Cotard's syndrome, image and object are disentangled, but the intrusion of the object does not affect the images of other human beings, as it is the case in false recognition syndromes; Cotard's syndrome involves another aspect of one's relationship to what Lacan calls the Other, namely the relationship to one's own body orifices.

therefore provides a precise illustration of the effects triggered by the failure of the symbolic operation we call “naming” in psychosis. Naming does not only allow us to designate something using a symbol; it also identifies the one who names in a symbolic — i.e., differential — mode. In the Fregoli syndrome, naming appears to consist of only one name, which identifies the object; the object is equivalent to this one name: it is always the same. One might argue that the psychotic failure of specular knowledge (Lacan, 1966/2002a) results, in various forms, in an impossibility to stabilize the apprehension of reality in specific representations. This produces multiple instances of doubling, which we also notice in thought-echo (Séglas, 1895) and in other elementary phenomena of mental automatism (de Clérambault, 1987).

Discussion

Since the Fregoli and Capgras syndromes were first described, various psychoanalytical interpretations have been proposed (Capgras and Carette, 1924; de Pauw, 1994). Ringenbach (1986) and Porge (1986) have commented on the real, imaginary, and symbolic aspects of Capgras syndrome, which they consider a disjunction between the body and the image or an abolition of the *méconnaissance* linked to specular identification. However, most psychological interpretations focused on the theme of Oedipal conflict. For example, Lykouras, Typaldou, Gournellis, Vaslamatzis, and Christodoulou (2002) discuss a possible psychological susceptibility in a patient who presented with both Fregoli and Capgras syndromes:

It is possible that organic deficits affecting the sense of familiarity were combined with a preexisting ambivalence towards the object, activating the defence mechanisms of denial, projection and pathological splitting of the internalized object representations. By “inventing” the bad double of her father the patient’s negative feelings towards her father, through a mechanism of pathological splitting and projection, were directed towards, the double sparing the father and thus avoiding guilt [...] Her love was also projected to the outer world through her Fregoli syndrome, which partially serves to an incestuous wish fulfilment. (p. 235)

The current literature on misidentification syndromes is mainly devoted to the cognitive disorders observed in these syndromes. For example, Christodoulou (see Papageorgiou, Ventouras, Lykouras, Uzunoglu, and Christodoulou, 2003) emphasizes the deficiencies in working memory and allocation of attentional resources currently described in delusional misidentification syndromes. These cognitive studies rest on two pillars: (1) likening the rare occurrence of delusional false recognition after right hemispheric brain lesions (Ellis, 1994) to the well-established connection between face-recognition deficit (prosopagnosia) and right hemisphere lesions (Buchtel, 2001); (2) equating people recognition

with face recognition. As a result, these studies are based upon cognitive or neuropsychological standardized tests, which are mainly devised for studying prosopagnosia (Walther, Federspiel, Horn, Wirth, Bianchi, Strik, and Müller, 2010), a neurological disorder whereby patients do not recognize faces, while exhibiting peripheral signs of affective covert recognition (e.g., skin conductance response). It is noticeable that this approach, which implies what Luauté calls a unitary concept (i.e., bringing together psychiatry and the cognitive sciences) of misidentification syndromes (Sansone, Luauté, Bidault, and Tiberghien, 1998), ultimately fails to account for a specific aspect of the delusional syndromes: the fact that patients *believe* in their false recognitions, which is precisely what makes these recognitions deluding (Coltheart, Menzies, and Sutton, 2010). Young (2009) claims that the delusion of subjective doubles arises because individuals immediately recognize their relatives, but no longer have any arousal response towards them. Indeed, Capgras patients do not exhibit any differential modification of skin conductance to familiar faces (Ellis and Lewis, 2001), a condition inverse of what is observed in prosopagnosia.⁵ According to Young (2009), this absence of a somatic emotional marker might be experienced by the Capgras patient, “not as a lack of affective response but as a loss,” and be perceived as “estrangement” when meeting his relatives. However, Breen, Caine, Coltheart, Hendy, and Roberts (2000) have reported cases of both neurological and psychiatric patients, in whom such a lack of affective response to familiar faces was not critical for the development of delusional misidentifications. Although Young considers that the sense of estrangement “should be an integral component within any explanation proffered,” he is forced to recognize that “as a final but nevertheless important point, the sense of estrangement experienced by the Capgras patient is not sufficient to produce the delusional belief that the person is an impostor; nor is it the whole story regarding the phenomenology underlying the condition” (2009, p. 637).

Such dead-ends in the current neuropsychological approach to delusional misidentification syndromes provide an illuminating example of one of the most common risks in analyzing clinical data, namely that we only recognize what we already know and miss what the patient actually says. Indeed, even though we have little access to the patient’s phenomenological experience, we can rely upon what the patient says. But we can only rely on it provided that we write down exactly what we hear. Writing down what the patient says is of prime importance, since writing “under the patient’s dictation” means giving up *recognizing*. By giving up recognizing, we mean giving up “understanding” what we think we hear, giving up the implicit hypothesis that what we hear is homogeneous with our field of consciousness, i.e., with the limited set of representations we are able

⁵For Young (2009), the affective response measured by the electrodermal response is supposed to result from the mental activity involved in the retrieval of person–identity information.

to recognize. It is probably because traditional psychiatrists wrote down their observations that they were able to identify the distinctive features of Fregoli and related syndromes, as Courbon and Fail (1927) did when they wrote: “Fregoli is a single being.” Such distinctive features were not immediately recognizable or “understandable.” But they could be identified because they were underpinned by material elements — the patients’ words, their grammatical logic, transcription, and reading — that were independent of recognition. These considerations are not unimportant. Indeed, what Fregoli syndrome teaches us regarding recognition is closely linked to this problematic, insofar as it permits us to separate out some fundamental features of what Lacan called specular knowledge (1966/2002a). A full development of these notions is to be found in Thibierge (2011).

Instead of looking for disturbances in cognitive processing, some authors have tried to identify the logic that might account for the patients’ erroneous beliefs. Following Cutting (1991), Margariti and Kontaxakis (2006) proposed to read delusional misidentifications syndromes not as disorders of the sense of familiarity, but as disorders of uniqueness. They “hypothesize that a common potential pathogenic factor underlying DMS could be a disorder of the sense of uniqueness. The ability to attribute uniqueness to the self and to surrounding people, objects, or places is a principal property of the adaptable mind that acts as a matrix for the identification process” (p. 261). This approach is interesting since it corresponds to the formulas we have proposed above: *the other is always the same* and *the same is always an other*. Our Lacanian reading gives us further insight into this disorder of uniqueness: as human beings, we can perceive objects and people as unique, but only insofar as the unique object that makes us desire is itself concealed. Only the identification with an image, which simultaneously neutralizes this object, gives us access to uniqueness — this is what Lacan’s concept of the mirror stage tries to demonstrate. In other words, the possibility of recognizing people and objects as unique is not a basic physiological ability, but is the result of a specific process, one that establishes a specific connection, a relationship between our image and our object. This connecting process makes it impossible for the image and the object to permanently coexist in our psychic reality. In delusional misidentification syndromes, this process happens to be ineffective and it is the presence — rather than absence — of the object *a* that causes all the characteristics of the image *i* described above. This is the reason why Fregoli and Capgras syndromes, as described by French traditional psychiatry, can make a significant contribution to the questions of recognition, both in practice and in theory. As in a chemical analysis, Fregoli and Capgras syndromes separate the two elements — *i* and *a* — which we can never find isolated in our clinical experience of neurosis.

A different version of the separation of *i* from *a* may be found in neurological disorders of body image. We have shown elsewhere that in these cases the pseudo-delusional personification of paralyzed limbs may be read as directly attesting the

intrusion of the object — which would otherwise be repressed — in the patient's psychic reality (see Morin, Thibierge, Bruguière, Pradat–Diehl, and Mazevet, 2005; Morin, Thibierge, and Perrigot). In these neurological cases, it is the organic failure of body image that makes neutralization of the object ineffective. The psychopathological interest of these cases is that they bring to the fore an element that, precisely because of repression, remains rather obscure and difficult to grasp in neurosis. The neurotic's searching is always driven by the same object. However, a neurotic person can never identify this object, except when anxiety indicates its incidence. This is the price to pay for the possibility of recognition.

In the final analysis, delusional misidentification syndromes do not seem to be named appropriately, at least from the psychoanalytical point of view. Indeed, we have shown that in Fregoli and Capgras syndromes, the object is not misidentified — as it is in neurosis — but *identified*, whereas it is the *recognition* of images that is disturbed. The presence of the unique object — which in this case is not lost — in the patient's psychic reality, results in a multiplicity of inconsistent images.

References

- Breen, N., Caine, D., Coltheart, M., Hendy, J., and Roberts, C. (2000). Towards an understanding of delusions of misidentification: Four case studies. *Mind and Language*, 15, 74–110.
- Buchtel, H.A. (2001). Left and right hemisphere contributions to physiognomic and verbal discrimination. *Neuropsychology*, 15, 597–606.
- Caine, D. (2009). Reflecting on mirror self-misrecognition. *Neuropsychanalysis*, 11, 211–226.
- Capgras, J., and Carette, P. (1924). Illusion des sosies et complexe d'Édipe. *Annales Médico-Psychologiques*, 12, 48–68.
- Capgras, J., and Reboul–Lachaux, J. (1923). L'illusion des 'sosies' dans un délire systématisé chronique. *Bulletin de la Société Clinique de Médecine Mentale*, 2, 6–16.
- de Clérambault, G. G. (1987). *Œuvres Psychiatriques*, Paris: Frénésie.
- Coltheart, M., Menzies, P., and Sutton, J. (2010). Abductive inference and delusional belief. *Cognitive Neuropsychiatry*, 15, 261–287.
- Courbon, P., and Fail, G. (1927). Syndrome d'illusion de Fregoli et schizophrénie. *Bulletin de la Société Clinique de Médecine Mentale*, 15, 121–125.
- Courbon, P., and Tusques, G. (1932). Illusions d'intérométamorphose et de charme. *Annales Médico-Psychologiques*, 14, 401–406.
- Cutting, J. (1991). Delusional misidentification and the role of the right hemisphere in the appreciation of identity. *British Journal of Psychiatry*, 159, 70–75.
- Czermak, M. (1986). Signification psychanalytique du syndrome de Cotard. In J. Clims (Ed.), *Pas-sions de l'objet* (pp. 205–236). Paris: Editions de l'Association Freudienne.
- Ellis, H.D. (1994). The role of the right hemisphere in the Capgras delusion. *Psychopathology*, 27, 177–185.
- Ellis, H. D., and Lewis, M. B. (2001). Capgras delusion: A window on face recognition. *Trends in Cognitive Science*, 5, 149–156.
- Freud, S. (1895). *Project for a scientific psychology* (standard edition, volume 1, pp. 283–397). London: Hogarth Press.
- Freud, S. (1915). *Repression* (standard edition, volume 14, pp. 146–158). London: Hogarth Press.
- Freud, S. (1919). *The uncanny* (standard edition, volume 17, pp. 219–256). London: Hogarth Press.
- Lacan, J. (1980). *De la psychose paranoïaque dans ses rapports avec la personnalité*. Paris: Seuil.
- Lacan, J. (2002a). On my antecedents. In B. Fink (Ed.), *Écrits* (pp. 51–57). New York: W.W. Norton and Company. (Originally published 1966)

- Lacan, J. (2002b). The mirror stage as formative of the function of the I as revealed in psychoanalytic experience. In B. Fink (Ed.), *Écrits* (pp. 75–81). New York: W.W. Norton and Company. (Originally published 1966)
- Lacan, J. (2002c). Presentation on psychical causality. In B. Fink (Ed.), *Écrits* (pp. 123–158). New York: W.W. Norton and Company. (Original published 1966)
- Lacan, J. (2002d). Remarks on Daniel Lagache's presentation: "Psychoanalysis and personality structure." In B. Fink (Ed.), *Écrits* (pp. 543–574). New York: W.W. Norton and Company. (Originally published 1966)
- Lacan, J. (2004). *Le séminaire. Livre X. L'angoisse*. Editions du Seuil. Paris: Champ Freudien.
- Leuret, F. (1834). *Fragments psychologiques sur la folie*. Paris: Crochard.
- Lykouras, L., Typaldou, M., Gournellis, R., Vaslamatzis, G., and Christodoulou, G.N. (2002). Coexistence of Capgras and Fregoli syndromes in a single patient. Clinical, neuroimaging and neuropsychological findings. *European Psychiatry*, 17, 234–235.
- Margariti, M.M., and Kontaxakis, V.P. (2006). Approaching delusional misidentification syndromes as a disorder of the sense of uniqueness. *Psychopathology*, 39, 261–268.
- Mojtabai, R. (1994). Fregoli syndrome. *Australian and New Zealand Journal of Psychiatry*, 28, 458–462.
- Mojtabai, R. (1998). Identifying misidentifications: A phenomenological study. *Psychopathology*, 31, 90–95.
- Morin, C., and Thibierge S. (2006). Body image in neurology and psychoanalysis: History and new developments. *Journal of Mind and Behavior*, 27, 301–318.
- Morin, C., Thibierge, S., Bruguière, P., Pradat-Diehl, P., and Mazevet, D. (2005). "Daughtersomatoparaphrenia" in women with right hemisphere syndrome: A psychoanalytical perspective on neurological body knowledge disorders. *Neuropsychanalysis*, 7, 171–184.
- Morin, C., Thibierge, S., and Perrigot, M. (2001). Brain, body image and language: A psychoanalytic perspective. *Journal of Mind and Behavior*, 22, 69–89.
- Papageorgiou, C., Ventouras, E., Lykouras, L., Uzunoglu, N., and Christodoulou, G.N. (2003). Psychophysiological evidence for altered information processing in delusional misidentification syndromes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 27, 365–372.
- de Pauw, K.W. (1994). Psychodynamic approaches to the Capgras delusion: A critical historical review. *Psychopathology*, 27, 154–160.
- Porge, E. (1986). Endosser son corps, *Littoral*, 21, 65–88.
- Ringenbach, A.M. (1986). Avatars du corps et de son enveloppe. *Littoral*, 21, 37–50.
- Sansone, S., Luauté, J.P., Bidault, E., and Tiberghien, G. (1998). Une conception neuropsychologique unitaire du syndrome de Capgras. *Annales Médico-Psychologiques*, 156, 433–444.
- Schreber, D.P. (2000). *Memoirs of my nervous illness*. New York: NYRB Classics.
- Séglas, J. (1895). *Leçons cliniques sur les maladies mentales et nerveuses*. Paris: Asselin and Houzeau.
- Thibierge, S. (2011). *Le nom, l'image, l'objet. Image du corps et reconnaissance*. Paris: Presses Universitaires de France.
- Thibierge, S., and Morin, C. (2010). The self and the subject. A psychoanalytic Lacanian perspective. *Neuropsychanalysis*, 12, 81–93.
- Walther, S., Federspiel, A., Horn, H., Wirth, M., Bianchi, P., Strik, W., and Müller, T.J. (2010). Performance during face processing differentiates schizophrenia patients with delusional misidentifications. *Psychopathology*, 43, 127–136.
- Young, G. (2009). In what sense 'familiar'? Examining experiential differences within pathologies of facial recognition. *Consciousness and Cognition*, 18, 628–638.

Does Functionalism Offer an Adequate Account of Cognitive Psychology?

James M. Stedman

University of Texas Health Science Center San Antonio

Thomas L. Spalding and Christina L. Gagné

University of Alberta

The majority of cognitive psychologists, when pressed, would hold for scientific realism as their philosophy of science and ontological position. However, it is functionalism which undergirds the ontology of scientific realism. Moreover, functionalists claim that cognitive psychology, including sensation, perception, memory, and the higher cognitive functions, can be fully accounted for by functionalism. The question is then: Is functionalism up to the task? Recently, Spalding and Gagné (2013) made the case that concept formation, a key element in all aspects of higher order cognition, can be better accounted for by an Aristotelian–Thomistic (A–T) meta-theory, and Stedman (2013) pointed out parallels between Aristotle’s model and current cognitive psychology. This essay argues that the A–T viewpoint is a better model for all elements of cognitive psychology.

Keywords: Aristotle, Aquinas, cognitive psychology, functionalism, dualism

Confidence in logical positivism and operationism waned in both philosophy and psychology during the 1960s. At about the same time, psychology experienced the cognitive revolution, which reinstated mental states and processes as central to theory building and explanation. As Levin (2013) points out in her review, cognitive psychologists turned to the functionalist theory of mind (Putnam, 1975) as a philosophical underpinning for all aspects of cognition. This interchange produced psycho-functionalism, as described by Levin:

A second strain of functionalism, psycho-functionalism, derives primarily from reflection upon the goals and methodology of “cognitive” psychological theories. In contrast to the behaviorists’ insistence that the laws of psychology appeal only to

behavioral dispositions, cognitive psychologists argue that the best empirical theories of behavior take it [behavior] to be the result of a complex of mental states and processes, introduced and individuated in terms of the roles they play in producing the behavior to be explained. . . . All versions of functionalism, however, can be regarded as characterizing mental states in terms of their roles in some psychological theory or other. (p.7)

Hence, as Levin sees it, functionalism in all its aspects is intertwined with cognitive psychology at both the philosophical and psychological levels.

The purpose of this essay is as follows: (a) to describe the philosophical roots of functionalism; (b) to present an overview of functionalism as it currently operates in cognitive psychology; (c) to review objections to functionalism; and (d) to examine an alternative account, the Aristotelian–Thomistic (A–T) model, recently proposed in cognitive psychology itself (Spalding and Gagné, 2013; Spalding, Stedman, Hancock, and Gagné, 2014) and within the philosophy of mind (Feser, 2006, 2014; Madden, 2013).

Analytic Philosophy and Functionalism

Analytic philosophy, spawned in Britain as a reaction to idealism, dominates English-speaking philosophy to the present time. As Preston (2007) demonstrates in his review of the history of analytic philosophy, the movement was initiated by Russell and Moore, refined as logical atomism by Russell and Wittgenstein (see Wittgenstein, 1922), elaborated by the Vienna Circle as logical positivism (Ayers, 1952), seriously questioned by Quine (1951), and later reinvented by Wittgenstein (1953). Preston characterizes contemporary analytic philosophy, the home of the philosophy of mind, as eclectic and interested in limited metaphysical problems, as still grounded in language analysis and semantics, and as interested in the kinds of thought experiments often used by philosophers of mind. Although current philosophers of mind are split along property dualism (Chalmers, 1996, 2010; Nagel, 1974) and materialist lines (Churchland, 1986; Lewis, 1966), all agree that cognitive processes, from sensation and memory through all of the higher order phenomena of thinking, reasoning, categorization, planning, etc., can be explained by the doctrine of functionalism. In fact, Chalmers (1996), though a property dualist, asserted this about functionalist cognitive models:

Cognitive models are well suited to explaining psychological aspects of consciousness. There is no vast metaphysical problem in the idea that a physical system should be able to introspect its internal states, or that it should be able to deal rationally with information from its environment, or that it should be able to focus its attention first in one place and then in the next. It is clear enough that an appropriate functional account should be able to explain these abilities, even if discovering the correct account takes decades or centuries. (p. 31)

In her review, Levin (2013) points out that functionalism has antecedents (Ryle, 1949; Turing, 1950; Wittgenstein, 1953) but emerged as a definitive philosophical position in the last 35 years of the twentieth century. Feser (2006, ch. 3) provides a brief and readable explanation of how functionalism developed in response to problems with other strictly materialist philosophies, such as philosophical behaviorism and identity theory. Several major strains of functionalism developed: machine functionalism, psycho-functionalism, and analytic functionalism. So, far from being monolithic, functionalism itself is divided, with arguments in favor of and attacking the various strains.

Machine functionalism (Putnam, 1975) was the first of the three developed and appeared to answer a number of problems with behaviorism, for example, critiques pointing out that behaviorists seemed to undervalue internal mental states, such as beliefs and desires, when those constructs contributed implicitly to their theories. Machine functionalism postulated that any "mind" can be regarded as a finite digital computer, one that receives inputs (1), while in a certain state (S1), goes into other states (Sx), and produces output (0). The person with a mind is viewed as a probabilistic automaton, so for each state and sets of inputs, the machine (mind) will enter a subsequent state and produce output according to certain probabilities. These internal mental states were called representations, and the nature of these representations was and is disputed. Over time, early machine functionalism's ties to a "machine table" or program came to be seen as inadequate. However, the basic idea involving inputs, lawful interactions of internal states, and final outputs has been retained in later functionalist approaches.

Psycho-functionalism is closely tied to the emergence of cognitive psychology in the late 1960s and maintains that mental states and processes are entities (constructs) that are defined by the role they play in cognitive psychological theories. They may be tied to brain structures and processes but this is not a requirement. However, there does seem to be a trend toward attempting to ground these constructs in neuroscience (see Stedman, Hancock and Sweetman, 2009). These constructs can include mental states and processes easily identified with common sense (folk psychology) or can go beyond common sense to incorporate more refined constructs identified by laboratory findings, thus replacing folk psychology constructs.

Analytic functionalism, developed by Lewis (1966) and Armstrong (1968), asserts that all mental states, such as pain, hunger, belief, desire, consciousness and so on, are constituted by their functional role, that is, to quote Chalmers (1996): "On this view, a mental state is defined wholly by its causal role: that is in terms of the stimulation that tends to produce it, the kind of behavior it tends to produce, and the way it interacts with other mental states" (p. 14).

Advocates of functionalism hold for the following: (a) functionalism offers a comprehensive ontological account either of all mental states or at least the psychological portion of mental states but not the phenomenal (Chalmers, 1996,

2010); (b) some mental states, according to Chalmers and others, are primarily psychological and are fully accounted for by functionalist ontology, including learning, memory, categorization, perception, and higher-order cognitive processes; (c) some mental states, such as belief, desire, and hope, are referred to as intentional mental states in that they refer to something about the world. Many functionalists regard these states as primarily psychological and, hence, fully accounted for by functionalism. Chalmers (1996) views mental states as mixed. However, whatever in these states is psychological, he asserts can be accounted for by a functionalist ontology.

As mentioned above, the models of cognitive psychology rest on functionalism as a philosophical foundation. Psycho-functionalism may carry a slightly different connotation, but it is really a variety of analytic functionalism. Hence, psycho-functionalism is subject to all the ontological strengths and weaknesses of analytic functionalism.

Functionalism in Current Cognitive Psychology

Contemporary cognitive psychology theories are grounded in psycho-functionalism, expressed as models and/or mechanisms: for example, perceptual binding (John, 2002), working memory (Baddeley and Hitch, 1974), category formation (Smith, Patalano, and Jonides, 1998), and so forth. Some theories incorporate brain structures and processes and others do not. All postulate multiple, interacting mental states and processes that play definite roles in the theory. All are grounded in stimuli at the beginning and behavioral outcome at the end. Various psycho-functional cognitive theories compete, and their truth claims are established by empirical observations.

For example, consider the exemplar model of concept formation. The exemplar model claims that category (concept) formation occurs when people compare new information to exemplars stored in memory. This version of concept formation states that exemplars are learned through repeated presentations and naming of category members and the repeated naming allows the pairing of a common name with a set of exemplars, which in turn allows generalization over those exemplars when the name (or other similar cue) is presented. The exemplar model requires a number of psychological constructs: sensation, perception, learning (of exemplars), many constructs in the area of memory and recall, some mechanism accounting for comparing new stimuli to exemplars, an account of language to perform the response. Because cognitive psychology is increasingly linked to neuroscience, interactions with brain structures must also be factored in. Similar examples could be offered for all content areas of current cognitive psychology, and psycho-functionalism is expected to serve as the ontological and epistemological underpinning for all of cognitive psychology.

Problems with Psycho-Functionalism and Functionalism in General

As mentioned earlier, both property dualist and materialist philosophers of mind agree that functionalism is sufficient to account for all the "psychological" components of cognitive psychology. However, functionalism, as a philosophical position, has been challenged by philosophers of mind almost from its inception. Two of these objections will be presented in some depth.

A general objection to functionalism, well known in the philosophy of mind, is the "damn/darn problem," brought forth by Block and Fodor (1972). Their objection lies within the broader theory of mental holism, which claims that the meaning of a belief (or a sentence expressing that belief) is determined by its place in a network of beliefs (or sentences) making up a whole theory or even a group of theories. Block and Fodor point out that a functional account of mental states must take into consideration any difference in stimuli or responses. Joe smashes his finger in the door and says, "Darn"; Clyde smashes his finger in the door and says, "Damn." We have two equivalent stimuli but different responses. However, functionalists claim that outputs are related to all or many of the agent's internal mental states, so two people who have pain but produce different outputs must share little, if any, common mental states. But this conclusion appears absurd. Hence, Block and Fodor believe that functionalism leads to an extreme and undefendable version of mental holism.

This critique is also important in that one of the presumed advantages of functionalism was that it would allow one to avoid the type/token problems that arose in identity theories. In particular, a problem with identity theories is that, if the mind/thought is taken to be identical to the brain/brain state, then there must be lawful relations between given brain states and thoughts, but there seem clearly to be logical relations among thoughts, but only physically (efficiently) causal relations among brain states. This difference calls into question rather strongly the whole notion of identity between brain states and types of thought. Functionalism works around this problem by defining states with respect to their functional (including logical) relations to other states. However, if, as Block and Fodor (1972) claim, differences in response (including between individual tokens of a particular type of thought) dictate large differences in mental states, the mental holism implied by this removes this purported advantage of functionalism. In short, functionalism does not avoid a similar kind of problem as that identified in identity theories.

A second, and perhaps more damaging, critique asserts that functionalism, as an ontological claim, is trivial, that is, that functionalism's internal structures (constructs), anchored by stimulus inputs and behavioral outputs, are not unique but can be present in many complex and less complex systems. This criticism was stated earlier by Putnam (1988), Searle (1990) and others and has been reworked recently by Godfrey-Smith (2008). These arguments all point out that

the functionalist realization of a complex mental system is present when the set of mental states maps or corresponds to physical states of the system. These systems are known as combinational state automatons (CSA) and can occur in a large number of formats, including a properly manipulated bucket of water! Godfrey-Smith (2008) asserts the following:

If a normal human's functional organization over some interval is represented by a CSA, then our designer could build a transducer device that perturbs the bucket of water in specific ways in response to every possible sequence of inputs that a human might receive, and another transducer device that maps the water's responses to appropriate behaviors. So a bucket of sea water could act as control system for a humanoid robot, provided that our designer was extraordinarily knowledgeable about the object's contingency tree and skilled in the building of input and (especially) output transducer devices. (p. 23)

The triviality argument, as mentioned, is very complex and only this brief sketch will be presented here. It is important to recognize, however, that the import of the critique goes beyond the fact that it seems to lead to strange outcomes in which a bucket of water might have a mind. In particular, the issue is that if these kinds of arguments go through, then there is no guarantee, in functionalist terms, that any mental or intentional state actually bears the meanings that we ordinarily assign to them (see, e.g., Madden, 2013, ch. 5). In sum, there is a real possibility that a consistently functionalist account of thought collapses upon close inspection (see also Feser, 2006). For example, functionalist treatments of intentionality tend to lead to a denial of, or eliminative reduction of, intentionality. But the functionalist treatment itself is presented and argued for in intentional terms. Similar problems arise in functionalist treatments of logical reason and other topics.

Other general objections to functionalism include how to characterize the inputs and outputs of a functional system (Block, 1990), problems in accounting for what appear to be the causal effects of our mental states (Kim, 1989), and introspective belief (Armstrong, 1968). All of these objections and more have been put forward, many from functionalism's early days. In summary, functionalism as a philosophical ontology has met with serious objections. It should be noted that these objections apply to all functionalist positions, including psychofunctionalism and analytic functionalism.

Within psychology, functionalism has failed in the task of theory building and theory discrimination. A recent example of this failure was presented by Spalding and Gagné (2013). They pointed out that none of the three predominant models of concept formation (exemplar, prototype, and theory-theory) has emerged as superior, and all have approximately the same amount of empirical support (this problem occurs in many areas of psychology). In fact, they note that recent studies of psychological essentialism, generics, K-properties, and perceptual

symbol systems/embodied concepts challenge the validity of all three probabilistic models mentioned above.

It should be noted that all three theories of concept formation meet the criteria demanded by functionalism. All commence with stimulus conditions, postulate interacting internal cognitive constructs, and produce behavioral outcomes. However, years of research have not led any of the three theories to dominance. In fact, Spalding and Gagné (2013) made this comment about the current state of affairs:

Given the diversity of representations and processing systems suggested by recent research results, it is unclear whether there can be a theory of concepts at all (e.g., Machery, 2009). How can concepts be both essentialist and yet not involve necessary and sufficient features? . . . How can they be both sensory-based and abstract/universal? (p. 71)

Hence, we see that this level of theory building and theory discrimination has failed. Of course, it is possible that there will be a new, functionalist-based, theory that will account for human concepts, though Machery (2009) makes a strong case that this is unlikely. However, this conspicuous failure can also be taken to suggest that a more fundamental re-thinking of concepts is in order, and, since failure of theory discrimination is common in most areas of cognitive psychology, perhaps functionalism is simply not up to the task of theory building.

The A–T Alternative

As mentioned in our statement of purpose, both philosophers of mind and psychologists have recently proposed Aristotelian–Thomistic alternatives to functionalism. We will consider philosophers of mind first. Madden (2013) has written a thorough review of the major approaches to the philosophy of mind, including functionalism. In fact, he is rather enthusiastic about functionalism, as noted in the following quote:

Functionalism is a powerful theory. It seems to provide an account of psychological states without even a hint of anything left to be explained by supposed nonphysical states, while at the same time allowing for mental causation without raising problems of mind–body interaction. . . . It is, whatever its vices, a very good idea. (p. 131)

Although Madden concedes that functionalism has many merits, he argues that its shortcomings outweigh its merits as a philosophy of mind. His arguments against functionalism are those covered in the previous section, plus issues regarding intentionality, beliefs, and thoughts in general.

As an alternative, Madden proposes Aristotelian–Thomistic hylomorphic theory as applied to human (and animal) sensory and cognitive processes, as the best solution to the observed phenomena of mind. His argument for this position is indeed complex and is marshaled against materialist positions, property dualism, and substance dualism; all fail as explanatory systems for mind. For the purposes of this essay, we will not present any further detail regarding these purely philosophical arguments raging among the philosophers of mind but will leave it to the reader to investigate the arguments in the works cited below.¹

Returning to our primary question regarding functionalism’s adequacy as an account of cognitive psychology, we have seen that functionalism has deficits as an ontological system. These deficits also apply to functionalism as an underpinning for cognitive psychology. Recognizing problems with functionalism and particularly incoherence in cognitive psychology’s account of concept formation led Spalding and Gagné (2013) to propose an Aristotelian–Thomistic alternative model to account for concept formation. At about the same time, Stedman (2013) published a paper demonstrating parallels between modern cognitive psychology and neuroscience and Aristotle’s hylomorphic theory. He showed that these parallels exist at all areas of interest to cognitive psychology and neuroscience, including sensation, perception, memory, and higher-order cognitive functions, such as concept formation, theory of mind, reasoning, and so forth.

The details of the A–T model have been elaborated in three recent publications (Spalding and Gagné, 2013; Spalding, Stedman, Hancock, and Gagné, 2014; Stedman, 2013) and so will be outlined here only briefly. The A–T framework commences with sensory information regarding objects in the environment. This information is organized by the “internal senses,” including the common sense (*sensus communis*), which receives and arranges all sense data, the phantasm, which retains the sense data, the imagination, which combines and reassembles images from the phantasm, and the memory, which retains sensory level images for later use. The intellect, by the process of abstraction, then acquires the universal form of the object. The A–T model calls for a second movement. For a concept to be finalized, the universal, held in mind, must be predicated. In this process, there is movement from the universal back down the internal senses, the phantasm in particular. This act, known as the existential

¹As mentioned in our statement of purpose, both philosophers of mind and psychologists have proposed an Aristotelian–Thomistic (A–T) account as an alternative to functionalism. Readable recent descriptions of A–T approaches to metaphysics (Feser, 2014), philosophy of mind (Feser, 2006; Madden, 2013), and other specific relevant topics, such as induction (Groarke, 2009) or essences (Oderberg, 2007), as well as overall systematic descriptions of Thomistic philosophy (e.g., Feser, 2009; Stump, 2003), and older descriptions specifically relating to the application of A–T ideas to philosophical and experimental psychology (e.g., Brennan, 1941; Maher, 1909), are available. Feser (2014) is particularly interesting in the current context as he takes a compare/contrast approach with many recent developments in analytic philosophy.

judgment, affirms the existence of the particular, as in "This man (universal) is Joe (particular man)."

To date, the most detailed presentation of the A–T model as an alternative to functionalism is that of Spalding and Gagné (2013) with regard to concept formation. Spalding and Gagné are careful to distinguish between the A–T and the classical cognitive psychology understanding of concepts at each step in their presentation of the A–T model. They do so in order to demonstrate that rejection of the classical cognitive psychology understanding of the concept has no implications for the A–T model. They summarize three crucial differences:

1. In the A–T model, concepts are "essences shared by all members of the category" and involve external and internal sense involvement of particulars at the outset and at the return to particulars, whereas as cognitive psychology's classical view considers the concept to be only a definition, composed of a bundle of properties.

2. In cognitive psychology's classical view, concepts are definitions of necessary and sufficient features and this view cannot deal effectively with exceptions, whereas the A–T model does not view concepts as constituted by necessary and sufficient bundles of features but by the abstracted essence, and the A–T model manages exceptions by the theory of privation or lack of actuation of essential characteristics in a particular individual.

3. In cognitive psychology's classical view, what "nests" in the concept are bundles of necessary and sufficient features; whereas in the A–T model what nests are capacities of the essence. In fact, the A–T model of concepts reflects the full A–T metaphysics: act and potency, substance and accidents, four-cause understanding, and the faculty psychology of Aristotle and Aquinas, a much more elaborate ontology than a bundle of features gathered by sense observation.

Having differentiated the A–T view from classical cognitive psychology, Spalding and Gagné next consider the current probabilistic models of concept formation. They mention three: (a) exemplars, (b) prototypes, and (c) theory–theories. The exemplar model states that concepts consist of the representations of the individual members of the categories covered by the concepts. The prototype model claims that concepts consist of single summary representations that, in one way or another, are summaries of the whole set of members of the categories covered by the concept. The theory–theories model asserts that concepts consist of explanatory "theories" that use existing background knowledge to identify and explain the categories covered by the concept. Although there is overlap among these models, all have somewhat different views regarding the nature of these probabilistic representations; and all three models have strong support from empirical research. Hence, no particular model has been able to establish itself as the superior.

Spalding and Gagné analyzed problems with each model. Regarding the prototype model, they point out that the model's primary advantage relative to

the classical view is that it includes features beyond the necessary and sufficient, and furthermore, that those features that are more frequent tend to be more important to the concept. They then note that the A–T view makes distinctions between features that are always true versus those that are there on a less frequent basis (basically the distinction between substance and accidents) and, furthermore, that the frequency effects identified in the prototype model result from the A–T view's use of sensory-based representations (i.e., phantasms) that contribute to the concept. Spalding and Gagné claim that this flexibility solves the problem of prototype match. However, the prototype model has great difficulty explaining the effects of particular individual category members, as the concept is only the prototype.

The strength of the exemplar model, compared to the prototype model, is that it provides a natural way of understanding how individual instances of a category can come to exert particular influence in the use of a concept. Spalding and Gagné believe that the A–T view also solves this issue through the phantasm model of the internal senses. In the A–T view, a particular exemplar or combinations of exemplars from memory (i.e., phantasms) must always be recruited when processing a particular category member. Hence exemplar-specific effects are expected in the A–T view. However, the exemplar model has serious difficulties in explaining how one can reason about a concept as a whole. For example, if one learns that whales are mammals, how does one incorporate this into the concept of mammals? Must one retrieve each stored exemplar of whale and attach the property? If there is some one representation where one can attach the new property, then how is that representation related to the supposedly diffuse representation of the concept via the exemplars? The A–T view solves this difficulty by predicating the property to the whole category via the intellectual concept.

The A–T view solves the concern about background knowledge as that affects concept formation. In particular, concepts seem to be much more than simply a statistical accumulation of features. We have many ideas about why certain features might co-occur, for example. In the A–T view, causal analysis of concepts by means of efficient, formal, material, and final causes (the hylomorphic theory) is parallel to modern theory–theories formulations of concepts and generalizations from those concepts. A weakness of the theory–theory, though, is that there is no obvious explanation for how the theory–theory, as a representation of the concept, also accounts for the statistical effects that gave rise to the probabilistic understanding of concepts in the first place. In the A–T view, on the other hand, the relation of the reasoning about causes to the understanding of the concept (and the category members captured in the phantasms) is spelled out. Spalding and Gagné (2013, p. 82) summarize their discussion of the three main theories of concepts in the following way, "The A–T view combines elements of each of the three main views that have developed within the probabilistic consensus

on concepts. Thus, from the A–T view, it is not surprising that there should be empirical evidence in favor of each of these perspectives."

Finally, Spalding and Gagné note that more recent research on psychological essentialism, K-properties, and perceptual symbol systems/embodied concepts challenge the notion that concepts are simply structures consisting of statistically aggregated features defining the concept. Recent work on each of these topics points to something more than frequentist (i.e., statistical) explanations of conceptual structure, and each has strong links back to the A–T claims about concepts.

Psychological essentialism (e.g., Gelman 2003, 2004) refers to the fact that people bring to bear beliefs about the cause and essential nature of the category covered by the concept, and that features playing a role in that belief are more important to the concept than they should be, based on statistical considerations. Research on K-Properties (e.g., Prasada and Dillingham, 2006, 2009) is closely related to psychological essentialism (and very closely related to the notion of essential vs. accidental properties in the A–T view). Again, K-Properties, those that are directly related to the "kind of thing" the concept covers, are more important than other, statistically equivalent properties. Finally, most concepts research has assumed, as a kind of default, that concepts are amodal. That is, both in terms of the conceptual representations themselves and in the representations of their features, concepts are not specifically linked to or grounded in sensory information. Yet, recent research (e.g., Barsalou, 1999) has shown important effects suggesting modality-specific information playing an important role. Again, this recent work harks back to the close linkage of the concept to sensory information, and the phantasms, in particular.

In sum, Spalding and Gagné (2013) surveyed several varieties of concepts research and showed that the results, though seemingly completely incompatible from the viewpoint of modern concept theories, all appear to be at least compatible with, and in most cases directly expected by, the A–T view of concepts. On the basis of this strong correspondence between the A–T view and the empirical results, Spalding and Gagné suggest a serious investigation into the A–T view as a possible high-level framework within which the scientific investigation of the psychology of concepts can proceed. Most importantly for the present purposes, Spalding and Gagné's presentation of the A–T view should be taken as a kind of existence proof that an A–T approach is compatible with modern scientific psychological research (see also Gagné, Spalding, and Kostelecky, in press, and Spalding and Gagné, 2015, for examples applying an A–T approach to a specific experimental research program within the field of concepts research).

Even more recently, Spalding, Stedman, Hancock, and Gagné (2014) examined the problem of intentionality in concept formation. They pointed out that cognitive psychology currently counts on functionalism, specifically psycho-functionalism, to explain the process of concept formation, including intentionality (what the concept is about). They note that the psycho-functionalist model acknowledges

the stimulus as a starting point; however, the primary explanatory focus of the functionalist paradigm is on the interactions of internal cognitive mechanisms and the behavioral outcomes:

However, psycho-functionalism's focus of explanation is on the mechanisms mentioned above and their interaction. The response element is more important than the stimulus because it is the empirical demonstration that the model has predicted correctly. Hence, with regard to intentionality, the "what the concept is about," the psycho-functionalist epistemological/ontological account has little to say. (p. 251)

Spalding et al. (2014) then turn to the A–T model and demonstrate that intentionality is totally achieved via the existential judgment cited above. Recall that in the existential judgment, there is a movement from the universal, back down through the internal senses, to affirm the existence of a particular, a "what the concept is about." For example, "The man (universal) is my friend, Joe (the existent person)." Again, the A–T model offers a better explanation than the functionalist model because of a definitive connection to the stimulus.

In sum, we have reviewed the philosophical origins of psycho-functionalism, discussed psycho-functionalism as the current philosophical underpinning of cognitive psychology, and pointed out its shortcomings at both the philosophical level and at the level of psychological theory building. We have proposed the A–T model as a positive alternative. In the next section, we will consider some objections to the A–T model.

Objections to the A–T Model

We are asserting that the A–T approach deserves a serious look by cognitive psychology, and we have argued that functionalism, the current meta-theory underpinning cognitive psychology, cannot withstand serious scrutiny. Thus, it is important to at least consider some of the potential objections to the A–T view.

One potential objection to the A–T view has been addressed by Spalding and Gagné (2013), namely, that the A–T view was ruled out by work on the so-called classical view of concepts in the 1970's and 1980's. It is worth noting that several recent presentations of A–T ideas have done something similar with respect to A–T philosophy more generally (see particularly Feser, 2008, 2009; Oderberg, 2007). This is quite important, as there is a fairly common, though unjustified, belief (a) that A–T philosophy has been shown to be wrong in some deep sense, and (b) that A–T philosophy is somehow inconsistent with modern science. But these claims are mostly the result of a kind of pious myth of progress and scientific revolution that disproved a bad philosophy and replaced it with a much better philosophy (see, e.g., Brown, 2006; Burt, 1925). Indeed, Feser (2014) explicitly argues that the A–T approach provides a superior philosophical

underpinning for science in general (see, particularly, pp. 9–25). In addition, Spalding and Gagné have argued that the A–T philosophical approach can serve as a meta-theory for concept formation, and, by extension, other areas of interest in cognitive psychology, including sensation, perception, memory, and the higher-order cognitive processes (see also, Stedman, 2013). Thus, there is no apparent fundamental incompatibility with the science of psychology.

A major concern some may have is that the A–T model reintroduces the Cartesian mind–body interaction problem full bore. This issue of substance dualism was covered in a previous essay (Spalding, Stedman, Hancock, and Gagné, 2014) but is so important it is worth very briefly restating: the mind–body distinction of Descartes and the A–T form–matter distinction are entirely different kinds of distinctions. There are (at least) two critical differences between this Cartesian understanding and the A–T view. First, the A–T view is very clear that mind and body are not two separate substances. Instead, the person is one substance made up of form (soul) and matter. The A–T view is clear that "primary substances" are ordinary individual things, all of which are made of form and matter, so this is not something unique about humans. Second, the Cartesian interaction problem arises due not just to Descartes' view of the soul but also to his mechanistic view of matter, such that only efficient causes are to be admitted. Thus, the soul must have some way of acting as the efficient cause of bodily actions and effects. Indeed, it is the adoption of Descartes' view of matter that necessitates a separate "thinking substance" and hence creates the problem of how that immaterial substance can affect the material "extended substance." However, if, with the A–T view, one understands the soul as a formal rather than efficient cause, the soul–body composite becomes a unity. Instead of the human person consisting of two distinct entities, whose interaction becomes a puzzle, the person is a unified entity, consisting of soul and body. The soul (the form of the body) actualizes the body to be species-specifically human. In light of this formal causality, the soul and body become two aspects of a single, unitary human being. Thus, the problem of interaction is avoided. There is one entity actuated by its form (soul or life principle), instead of two altogether different substances trying to interact. Rather than a substance dualism, the hylomorphic A–T approach is a kind of "uniformism" due to this distinct causal and metaphysical analysis.

One final point that might be made against adopting the A–T view is the following. Many psychologists will go about their work with little thought to its ontological underpinnings, and while we have shown that the A–T view would serve as a more satisfying underpinning, we suspect most psychologists will not regard ontological underpinnings as important. Our answer to this is twofold. First, a better philosophical underpinning for psychology is simply a better philosophical underpinning and should be preferred for that reason! But, second, most philosophical approaches since Descartes have tended to collapse into either a strictly materialist approach or into a purely idealist approach, both

of which have severe problems dealing with much of psychology. We would suggest that a similar collapse will occur over time (indeed, to a large extent has already occurred) within the discipline of psychology. Some areas move toward a pure, mechanistic materialism (e.g., many neuroscientists, medically-oriented clinicians, and even cognitive psychologists), while others move toward a more phenomenal approach (e.g., many people interested in personality, some clinicians, and others who reject purely mechanist materialism). Of course, if one is a true believer in either of these approaches, then one sees the other side as simply wrong-headed and perhaps does not regret this break. This requires, in some sense, the psychologist to dismiss not only the specific work done in the other tradition, but even to reject the interests in those areas entirely. An historic example of this is behaviorism's rejection of all mental phenomena. But many psychologists would likely wish for an ontology that breaks down this bifurcation, even as many psychologists rejected the behaviorist's rejection of the mental while still believing that the study of behavior and learning (for example) was a valid and important area for research. Indeed, one of the prime promises of functionalism (albeit a promise that we have argued is not actually fulfilled) is precisely to find a way to account for "mind things" without eliminating them. The A-T approach does not collapse into either materialist or idealist approaches, and potentially provides a philosophical underpinning to psychology that would allow psychology to remain an integrated discipline.

The A-T approach, by its adherence to a moderate realism and to hylomorphism, avoids both the pure materialist and pure idealist positions. It is particularly worth noting that the A-T approach has these advantages, not by inventing clever new ideas about the mind, but by consistently applying its own metaphysical concepts. For example, Madden (2013) contends:

As we turn to the implications of hylomorphism for the philosophy of mind, remember that the Aristotelian does not arrive at this view as an ad hoc attempt to gerrymander an account of nature around our commitments in the philosophy of mind. Rather, the Aristotelian takes this view in order to solve broader philosophical problems. (p. 251)

Conclusions

The major outlines of A-T and functionalism as ontological and epistemological systems have been presented here and in the recent publications mentioned above. We have also considered some of the philosophical strengths and weaknesses of both functionalism and the A-T model. While much thought has been devoted to philosophical critiques of both systems, we have not attempted to summarize that material in a few lines. We simply assert that, in a head to head comparison, the A-T hylomorphic model is better than functionalism as a scientific realist account of human cognition. Arguments for this assertion are found throughout

this essay, in the three recent publications cited (Spalding and Gagne, 2013; Spalding, Stedman, Hancock, and Gagne, 2014; Stedman, 2013). Functionalism claims to be an adequate ontology to explain all cognitive events from sensation to all higher-order cognition. However, as pointed out above, functionalism has been seriously challenged as an adequate general ontological theory. Spalding and Gagné (2013) and Stedman (2013) both make a reasonable case for A–T ontology as a meta-theory that provides at least as good a philosophical underpinning for cognitive psychology as functionalism does, in that (a) the A–T approach is compatible with the existing cognitive psychology research results, and (b) the A–T approach is capable of accounting for many aspects of the concepts literature that are deeply surprising if one takes a functionalist approach. Finally, we have shown that there are also philosophical reasons to prefer the A–T approach, and have shown that some of the common objections to such an approach are misguided, and often based on a misunderstanding of the approach.

References

- Armstrong, D. (1968). *A materialistic theory of the mind*. London: Routledge.
- Ayers, A. J. (1952). *Language, truth and logic*. New York: Dover.
- Baddeley, A. D., and Hitch, A. D. (1974). Working memory. In G. Bower (Ed), *The psychology of learning and motivation, Volume 8* (pp. 47–89). New York: Academic Press.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Block, N. (1990). Can the mind change the world? In G. Boolos (Ed.), *Meaning and method: Essays in honor of Hilary Putnam* (pp. 137–170). Cambridge: Cambridge University Press.
- Block, N., and Fodor, J. A. (1972). What psychological states are not. *Philosophical Review*, 81, 159–181.
- Brennan, R. E. (1941). *Thomistic psychology: A philosophic analysis of the nature of man*. New York: Macmillan.
- Brown, M. (2006). *Restoration of reason: The eclipse and recovery of truth, goodness, and beauty*. Grand Rapids, Michigan: Baker Academic.
- Burt, E. A. (1925). *The metaphysical foundations of modern physical science: A historical and critical essay*. London: Kegan Paul, Trench, Trubner, and Co.
- Chalmers, D. (1996). *The conscious mind in search of a fundamental theory*. New York: Oxford University Press.
- Chalmers, D. (2010). *The character of consciousness*. Oxford: Oxford University Press.
- Churchland, P. (1986). *Neurophilosophy*. Cambridge, Massachusetts: MIT Press.
- Feser, E. (2006). *Philosophy of mind: A beginner's guide*. Oxford: Oneworld.
- Feser, E. (2008). *Last superstition*. South Bend, Indiana : St. Augustine's Press.
- Feser, E. (2009). *Aquinas: A beginner's guide*. Oxford: Oneworld.
- Feser, E. (2014). *Scholastic metaphysics: A contemporary introduction* (Editiones Scholasticae: Volume 39). Heusenstamm, Germany: Editiones Scholasticae.
- Gagné, C. L., Spalding, T. L., and Kostelecky, M. (in press). Conceptual combination, property inclusion, and the Aristotelian–Thomistic view of concepts. In J. Hampton and Y. Winter (Eds.), *Compositional and concepts in linguistics and psychology. Language, cognition, and mind*. New York: Springer.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Science*, 8, 404–409.
- Godfrey-Smith, P. (2008). Triviality arguments against functionalism. *Philosophical Studies*, 145, 1–34.
- Groarke, L. (2009). *An Aristotelian account of induction. Creating something from nothing*. Montreal and Kingston: McGill-Queen's University Press.

- John, E. R. (2002). The neurophysics of consciousness. *Brain Research Reviews*, 39, 1–28.
- Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives*, 3, 77–108.
- Levin, J. (2013). Functionalism. *The Stanford encyclopedia of philosophy* (pp. 1–41). <http://plato.stanford.edu/entries/functionalism>
- Lewis, D. (1966). An argument for identity theory. *Journal of Philosophy*, 63, 17–25.
- Machery, E. (2009). *Doing without concepts*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195306880.001.0001
- Madden, J.D. (2013). *Mind, matter and nature: A Thomistic proposal for the philosophy of mind*. Washington, D.C.: Catholic University of America Press.
- Maher, M. (1909). *Psychology: Empirical and rational* (sixth edition). London: Longmans, Green, and Co.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Oderberg, D. S. (2007). *Real essentialism*. New York: Routledge.
- Prasada, S., and Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, 99, 73–112.
- Prasada, S., and Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33, 401–448.
- Preston, A. (2007). *Analytic philosophy: The history of an illusion*. London: Continuum International Publishing Group.
- Putnam, H. (1975). The nature of mental states. In H. Putnam, *Mind, language, and reality* (pp. 429–440). Cambridge: Cambridge University Press.
- Putnam, H. (1988). *Representation and reality*. Cambridge, Massachusetts: MIT Press.
- Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review*, 60, 20–43.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Searle, J. (1990). Is the brain a digital computer? *Proceedings and addresses of the American Philosophical Association* (Vol. 64, pp. 21–37). Newark, Delaware: American Philosophical Association.
- Smith, E. E., Patalano, A., and Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, 65, 167–196.
- Spalding, T., and Gagné, C. (2013). Concepts in Aristotle and Aquinas: Implications for current theoretical approaches. *Journal of Theoretical and Philosophical Psychology*, 3, 71–89.
- Spalding, T. L., and Gagné, C. L. (2015). Property attribution in combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 693–707.
- Spalding, T. L., Stedman, J. M., Hancock, C., and Gagné, C. L. (2014). Intentionality and the Aristotelian–Thomistic view of concepts. *Journal of Mind and Behavior*, 35, 245–262.
- Stedman, J. (2013). Aristotle and modern cognitive psychology and neuroscience: An analysis of similarities and differences. *Journal of Mind and Behavior*, 34, 121–132.
- Stedman, J., Hancock, C., and Sweetman, B. (2009). Cognitive psychology, neuroscience, and the problem of abstraction. *Contemporary Philosophy*, 29, 28–36.
- Stump, E. (2003). *Aquinas*. London: Routledge.
- Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus* [C. K. Ogden, Trans.] London: Routledge and Kagan Paul.
- Wittgenstein, L. (1953). *Philosophical investigations* [G. E. M. Anscombe, Trans.] Oxford: Blackwell.

Consciousness is Not a Physically Provable Property

Catherine M. Reason

London, United Kingdom

I present a logical proof that computing machines, and by extension physical systems, can never be certain if they possess conscious awareness. This implies that human consciousness is associated with a violation of energy conservation. I examine the significance that a particular interpretation of quantum mechanics, known as single mind Q, might have for the detection of such a violation. Finally I apply single mind Q to the problem of free will as it arises in some celebrated experiments by Libet.

Keywords: interactive dualism, hard problem of consciousness, Turing test

In 1995 Gilbert Caplain published a paper entitled “Is Consciousness a Computational Property?” in which he outlined an argument to the effect that no computing machine could ever be conscious. In his paper, Caplain pointed out that his argument was presented only in outline, and that some of the ideas presented required further work (Caplain 1995, 2000). In my opinion, Caplain’s argument is not, in fact, an argument that consciousness is not a computational property but rather something more subtle; it is an argument that no computing machine can ever, using purely computational processes, be certain if it is conscious.

To establish his argument, Caplain demonstrates an inconsistency between two principles: the principle of *reflexivity* and the principle of *cognitive separation*. Reflexivity is Caplain’s term for the capacity of conscious beings to know with certainty that they are conscious; cognitive separation can be expressed as the separation between some symbolic state in a computing machine, and the state of affairs which that state represents. Caplain argues that, if all computing machines are bound by the principle of cognitive separation, then the inconsistency between these two principles implies that no computing machine can ever be truly conscious, and hence conscious human beings cannot be computing

machines. This argument effectively applies Descartes' notion of the malicious genius to the internal states of a computing machine.

It seems to this author that Caplain's use of the term *reflexivity* does not conform to the usual philosophical usage, and so I shall use the term *self-certainty* instead. To avoid any ambiguity I shall define this term here:

Definition: Self-certainty is the capacity of at least some conscious beings to verify with certainty that they are conscious.

The detailed proof of Caplain's result that I am presenting here is substantially different from Caplain's in form, and attempts to minimize any dependence on philosophically ambiguous terms such as "knowledge" and "belief." However it relies on the same properties of consciousness and of machines. For the purposes of this argument, a computational process is operationally defined as any process which can be represented in the following form:

$$\text{Result} = P(\text{input})$$

where P is some computation. The exact form of P itself is irrelevant to this argument, so according to this definition a computational process is any computation which associates an input to an output. A computation here means simply any process which occurs in a computing machine. If the reader is concerned that this leaves the term "computing machine" undefined, then this may be taken to mean "some Turing machine," although this is not in fact a necessary stipulation.

In order to show that no computing machine can verify with certainty that it is conscious, one must first assume a computing machine M, all of whose computations are assumed to take the form above. At this point I shall also define the following Principle F (the functionalist principle):

"Every human mental process supervenes on some computational process."

This principle asserts, in effect, that human beings are computing machines of the same form as M. M is now presented with the task of proving that it is conscious. At this point two conditions must be noted:

1. M is given the task of proving that it is *certainly* conscious. Proofs that M may be conscious which depend on additional assumptions, or which fall within particular limits of confidence short of full certainty, are outside the scope of this argument and are not relevant to it.
2. "Conscious" in this context, does not necessarily mean "awake" or "self-conscious." It means only that some form of conscious experience is present, even if this is some altered state of consciousness such as a lucid dream. (It may seem odd to attribute such states to machines, but as it is impossible to assert, a priori,

what forms consciousness may take in computing machines, this possibility must be allowed for.)

At this point the reader should be careful to attend to the following operational definitions. Firstly I operationally define *certainty* as follows: M is certain of some proposition k if M is able to determine that k is certainly true. Other definitions of certainty — for example, subjective “feelings” of being certain — are not relevant to this argument. Secondly I operationally define *provable*, in statements of the form “proposition k is provable by M” as meaning: M is able to determine that k is certainly true. The reader should be careful not to confuse this operational definition with more familiar notions, for example those concerning the proof of theorems in formal systems.

M’s task can now be represented as a function or mapping from a domain E to a range X. E is a binary variable which represents the presence or absence of conscious experience and takes the following values:

E = 1 if conscious experience is present when the mapping is performed;

E = 0 if no conscious experience is present.

X is a binary variable which takes the following values:

X = YES if E = 1

X = NO if E = 0 or if the state of E cannot be ascertained.

The mapping therefore associates a state E, which represents the presence or absence of consciousness, with a state X which represents the answer to the question “Am I conscious?” This mapping is performed by a computation P which can be represented as follows:

$$X = P(E)$$

where X and E can now be thought of as states (or sets of states) in M. It is necessary also to make the following assumptions:

1. M can reason deductively (in particular, M must have deductive reasoning powers equivalent to those of a human being). It is not necessary to specify exactly what these powers are; merely that there is an equivalence between humans and M.
2. M is “honest” — that is, there are no systematic biases which prevent M from reasoning deductively in the domain in question. This is actually quite a difficult requirement to make precise. The best approach is to assert that there are no systematic biases which would make it impossible, even in principle, for M to follow classical rules of inference such as modus ponens.

I now define the following deductive argument which I shall call Argument A:

The reliability and accuracy of the computational process $X = P(E)$ depend critically on the reliability and accuracy of P (which is to say, how well P performs the mapping from E on to X). Consider some malformed computation BadP such that

$$X = \text{BadP}(E = 0) = \text{YES}$$

In such a case, M will conclude that it is conscious, but M 's conclusion will be neither accurate nor reliable. Therefore some further computation P' is required to establish the accuracy of P :

$$X' = P'(P)$$

where X' is YES if P is found to be accurate and NO otherwise. But what of the reliability and accuracy of P' ? Clearly this would necessitate some further computation P'' to establish the accuracy of P' — and so on, leading to an infinite regress. It follows that the reliability and accuracy of P can never be ascertained with certainty, and hence the value of E cannot be ascertained with certainty either. (One can paraphrase this by saying that, in any system which relies entirely on computations, the reliability and accuracy of any given computation can only be determined by applying another computation to it, and this process is obviously non-terminating.) It should be noted here that this argument applies even if $P = P'$ (that is, if P and P' are the same process) since it does not follow that $X = X'$. (As the input is different, the output can be different even if the function is the same.)

It follows from this that X cannot be guaranteed to be a reliable indicator of the value of E , nor can the value of any subsequent state, such as X' , render X ultimately reliable as an indicator of the value of E . In plain language this means that X , which represents M 's answer to the question "Am I conscious?," can never be relied upon to be a certainly correct answer to that question, so long as the value of X is determined by some computation. It is not possible, by means of any computation, to establish with certainty the value of E , and since M is a computing machine, M can never establish with certainty that it is conscious. This concludes the definition of Argument A.

It follows from Assumptions 1 and 2 that M can deduce A , and thereby deduce that M can never be certain if it is conscious. This rules out the possibility that M could be conscious, and arrive at the correct conclusion that it is conscious via faulty reasoning. Given our assumptions, it is simply impossible for M to be certain that it is conscious. It is important to note the two stages of this process. Argument A simply implies the potential *unreliability* of M (M may be accurate but it is impossible to establish this with certainty by means of any

computation). Assumptions 1 and 2 allow M to deduce A and thus deduce the *uncertainty* of M (M can show that it can never be certain of the accuracy of any of its computations). [Incidentally, it is not necessary for M to assume that it is a computing machine; it is sufficient for M to be unable to establish with certainty that it is not a computing machine.] This argument has a recursive character which may seem a little baffling at first sight, since the reader's brain is itself part of the argument! That is, I rely on the reader to appreciate the soundness of the deductive argument A. Once this is given, then Assumption 1 guarantees that M will also appreciate the soundness of A.

It is now apparent that M cannot possess self-certainty. But conscious human beings *do* possess self-certainty; it is possible for a conscious human being to know, with absolute certainty, that she is conscious (in the sense defined above in Condition 2). This implies that Principle F (which asserts that human beings are computing machines of the same form as M) must be wrong. It is in this sense that we can say that consciousness is not a computational property — or that if it is, it is attended by some other property or properties which are not themselves computational in nature. At this point it should be remembered that this proof applies only if M possesses deductive powers similar to those of a human being (Assumption 1). Conceivably if M did not possess such powers, then M could not deduce the argument A, and the proof of M's uncertainty would not apply; however in such a case, human beings could not be machines of the same form as M.

Expansion of the Computational Argument to Physical Processes

In the previous section, P was considered to be a computation mapping E on to X. However there is no reason to confine the definition of P in this way. P can instead be regarded as any *physical process* which performs the same mapping, and M can be regarded as a *physical system* rather than specifically a computing machine. To eliminate any confusion between mappings, computations, and physical processes, the relation between P and X can be rewritten to avoid any explicit mention of E:

$$X = O(P)$$

X is a binary variable as before, but P is now a physical process whose output O determines the value of X, where X is some state (or set of states) in a physical system. This formulation is intended to make it clear that physical processes which perform functions or mappings may not in any sense “look like” computations; in other words, they may not take the form of operations on data inputs. Once again, the reader may worry that the term “physical process” is effectively undefined. A physical process can therefore be operationally defined as any concrete entity in the real world which has the potential to evolve in time. This includes for example

collections of molecules, or computers running programs, but excludes abstract entities such as mathematical functions, or programs without implementations to run them. The output O of a physical process can be regarded as just the effect which that process has on the value of X . A *physical system* can be regarded as some set of physical processes.

It is now also necessary to change the Principle F to the following Principle F' (the physicalist principle):

"All human mental processes supervene on some physical process."

Argument A then proceeds much as before, except that the word "computation" in A is replaced by the word "process." Again one notes the possibility of physical processes BadP such that:

$$X = O(\text{BadP}) = \text{YES}$$

even when $E = 0$. This necessitates some physical process P' to ascertain the accuracy and reliability of P , and as before, this leads to an infinite regress.

This is all that is needed to show that, either consciousness is not a physical property, or it is attended by a property or properties which cannot themselves be physical. As before, Assumptions 1 and 2 imply that M can deduce the Argument A, and thereby establish that it can never be certain of being conscious. The upshot is that any physical system capable of reasoning honestly and which has deductive reasoning powers equivalent to those of a human being, would have to conclude that the question "Am I certainly conscious?" is effectively undecidable. Consciousness, therefore, is not a physically provable property.

How can this be? It is an inevitable consequence of the separation between the state of X and the process P by which the state of X is determined. This is analogous to Caplain's principle of cognitive separation. But it can readily be seen that it applies to *any* process P such that X is the output of P . In fact, even the qualifier "physical" is redundant; this argument applies to any sort of process *whatsoever* if the state of X is determined by the output of that process, rather than directly by E with no intervening process of any sort.¹

The reader may feel that this limitation on the capabilities of physical systems is too trivial to be worth mentioning. It simply means that human beings derive their certainty of being conscious not by any sort of mediating process, but by what in philosophy is called "acquaintance." However it has a serious consequence which has received virtually no attention within the academic literature. Principle F'

¹In fact it is not enough for E directly to determine X ; E must also directly determine that it is the case that E directly determines X , and do so in a way that a conscious subject can be certain is reliable — that is, not by means of any physical process which would be susceptible to Argument A.

implies that if M cannot be certain that it is conscious, then human beings cannot be certain that they are conscious either. Principle F' is therefore inconsistent with the property of self-certainty.² So — either Principle F' is wrong, or one of the other assumptions does not apply to human beings.

Assumption 1 cannot be discarded since by definition it must apply to human beings. Assumption 2 could be discarded but would leave one with the somewhat paradoxical situation that humans could be certain of being conscious only because their brains were incapable of honest reasoning (and hence were unreliable). Nonetheless, as we shall see later, there may be situations in which Assumption 2 could at least be modified, though to discard it entirely would be asking rather a lot of coincidence; it would in effect require a faulty system to produce, and produce reliably, the correct result via a series of fortuitous accidents. There could also be no way for humans to establish with certainty that the flaw in their reasoning was precisely that flaw required, for them to reach the correct answer to the question “Am I conscious?” This seems to leave one with no choice but to throw out Principle F'. Human mental processes, in other words, do not all supervene on physical processes.³

It is important to note that this conclusion applies not only to consciousness itself, but to some of the contents of consciousness as well. It also follows from Argument A that if human beings were exclusively physical systems, they could not be certain of the truth of the statement “I am reading this article”; indeed they could not even be certain of the truth of the statement “I believe I am reading this article.” One could even formulate Argument A in such a way that physical systems could not be certain of their own existence.

There is also an important difference between this conclusion concerning consciousness in physical systems, and the original, more restricted conclusion regarding computing machines. This is because even if human beings can be certain that they have conscious experience, it is still the case that physical systems — such as brains — cannot. This implies that when human beings ask themselves if they are conscious, either the evolution of their mental processes will diverge from the physical evolution of their brain-states in some drastic and irreversible manner; or their mental processes will force their brains to evolve in a manner which is inconsistent with their own physically determined behavior.

²Another way of looking at this is to say that knowledge or understanding by “acquaintance” is impossible in any physical system; or that if it is possible, it cannot influence the evolution of that system.

³One might think that allowing X to be identical with E might solve this problem — that is, by allowing X to be a state which is identical with consciousness itself. It is obviously possible to arrange things so that if E and X are identical, then it must be the case that E = 1 if X = YES. But to make use of this (and thus to be certain that X can be relied on) M must have some way of being certain that it is the case that E and X are identical. Since M is a physical system, any means of obtaining such a proof must supervene on some physical process, whereupon Argument A proceeds as before.

Such a violation of physically determined behavior should entail — at the very least — a violation of the principle of conservation of energy. Such a violation I shall henceforth refer to by the symbol χ (from *χιρραμιος*, a *cleft* or *gap*). The point of interest here is that χ should be empirically detectable. When human beings are asked to consider Argument A, and then decide if they are conscious, then — assuming all human beings are conscious, and know it — χ should be detectable within their brains.

Single-mind Q May Partially Conceal χ

I hope to examine the problems associated with the detection of χ in future work. However it is first necessary to examine a possibility which may make χ intrinsically undetectable, at least under certain conditions. This section will require a small diversion into quantum mechanics. In the most common interpretation of quantum mechanics (the Copenhagen interpretation) the physical state of a quantum system is represented by a vector in Hilbert space (Von Neumann, 1955). This state evolves deterministically according to the unitary dynamics of quantum mechanics (Barrett, 1999). Measurements are represented by applying an appropriate operator (in the form of a matrix) to the state vector, which produces a representation of the state vector in terms of some particular measurement basis. The physical states represented by the measurement basis are called eigenstates, since these are the states which result when the state vector is an eigenvector of the corresponding operator. Normally, however, the state vector will be a superposition of basis states, and on measurement this vector is assumed to “project” non-deterministically to an eigenstate of the measurement basis. This is the well-known “collapse” or “reduction” of the state vector.

The problem is that quantum mechanical theory does not provide any clear explanation of what constitutes a measurement. In order to circumvent this difficulty, attention has focused recently on so-called “no-collapse” interpretations, in which the physical state never collapses and superpositions persist indefinitely (see Barrett, 1999 for a review). However, this now presents us with another problem: how to account for the determinate nature of our experiences, which are always of single “classical” properties and never of superpositions of states. One approach to dealing with this is the single mind Q interpretation (Barrett, 1999). Single mind Q assumes some particular property Q, which evolves in such a way as to ensure that all our experiences are determinate. But in this approach, Q is regarded as a purely mental property — single mind Q, in other words, postulates a robust mind–body dualism. Q also functions to orchestrate or coordinate the experiences of different minds; without this, different minds would experience completely different and potentially unconnected realities.

The consequences of this for detecting χ are as follows. The process of neuroscientific inquiry can be regarded as the partitioning of a set U, which contains

every possible neural topography. Each element of U — that is, each neural topography — is a fully specified set of neurological properties (or as fully specified as quantum indeterminacy will allow). The term “neurological properties” is intended to refer to *all* brain properties, and not necessarily just positional ones. The partitioning of U will yield a subset which I shall call N . As neuroscientific inquiry advances, the set N would be expected to get smaller and smaller.⁴

However in a no-collapse theory, the physical state of the brain underlying U is assumed always to be a quantum state. It is important to be clear about what is going on here. The elements of U are not themselves quantum states. In fact in single mind Q , they are not really physical states at all. They are best understood as classical *appearances*; that is, they are descriptions of how neural topographies appear to the neuroscientists who are observing them. They are purely mental properties. (The determinate nature of these experiences is guaranteed by the determinate property Q , which is a property of the combined system of observer plus brain being studied.)

There are two ways in which U can be partitioned. First, as the physical system evolves, correlations will develop both between neurological properties and other neurological properties, and between neurological properties and properties in the environment. As this evolution occurs some elements of U will become inconsistent with the physical state. One can say that these elements are *partitioned out* of U , and not included in N . The second way U can be partitioned is via the process of quantum measurement; that is, the selection of an eigenstate for some observable. Since in single mind Q this sort of partitioning is always a mental process, the physical state remains unchanged after each partition. However the net effect of both types of process is to produce a subset N which is smaller than it was before.

There is here a potential loophole by which the effect χ might be partially concealed. Consider how the brain is normally thought to function; it is a physical system which instantiates what might be called *intelligent processes*. These are processes which enable the brain to respond to a wide range of environmental stimuli without requiring a separate programmed behavior for each stimulus. (Assumptions 1 and 2 may be regarded as an operational definition for such intelligent processes in human beings.) The understanding of these processes is

⁴The technically minded reader will have noticed that this is somewhat oversimplified. Although the classical requirement that neuroscientific inquiry is possible ensures that the subset N will reduce in size over time, quantum indeterminacy means that it will not do so smoothly; individual elements of U will “jump” in and out of N as N is refined. The reason for this apparent anomaly is that, in order to keep the representation simple, I have deliberately ignored the difference between static topographies — those defined at some precise instant of time — and dynamic topographies, which evolve over time. Neuroscientists who aim to understand the brain are typically interested in dynamic topographies. If one assumes that quantum mechanics plays no functional role in neural processing, then the dynamic topographies can be considered as evolving in essentially classical ways. In this case the quantum indeterminacy in the static topographies can be considered as noise and disregarded. From a neuroscientist’s perspective, the physical state can therefore be regarded as a set of classical topographies which is subsequently partitioned by measurement.

the business of the so-called “special” sciences, such as psychology and cognitive neuroscience. Intelligent processes are assumed to supervene on the physical processes which instantiate them.

Now consider the following thought experiment. Imagine an enormously powerful oracle, which is able to give accurate and meaningful answers to every question asked of it. Such an oracle would appear omniscient to all those by whom it was questioned. But consider that the actual number of questions such an oracle is likely to be asked in a finite period of time is probably a very small fraction of the number of questions which *could* be asked. If it were possible for the oracle to know in advance which questions would be asked, then the oracle could perhaps contrive to know the answers to just those questions and not trouble itself about those questions which no-one would ask. The oracle would still appear omniscient to all those who questioned it; but in practice it would be no such thing.

An analogous situation potentially exists in the relationship between intelligent processes and the physical processes which instantiate them. Of course no-one believes that intelligent processes are all-powerful, but they are very likely far more powerful than is needed to deal with the whole range of situations which arise within a given human lifetime. That is, intelligent processes are capable of dealing with many situations that never in fact arise. This is assumed to be necessary because no-one can predict what situations will actually arise within a human lifetime, even though most of them will never occur. But what if the actual state of the brain were indeterminate at the moment each novel environmental situation arose? In that case the conscious experience of each new environmental stimulus could be regarded as a further partitioning of N . If the actual state of the brain were indeterminate then the resulting partition would contain all neural topographies consistent with the correct response to that stimulus (except those which had previously been partitioned out of N). In most cases this would include all topographies which fully instantiated intelligent processes, but would also include many topographies in which intelligent processes were only partially instantiated (because these topographies would not yet have been partitioned out by measurement).

In the previous section it was shown that any physical system which fully instantiates the human capacity for deductive reasoning will be unable to conclude with certainty that it is conscious (or indeed that it has any other property). But this does not necessarily apply to systems which only *partially* instantiate human deductive reasoning. How does this work in practice? Successive neurological observations and conscious experiences will partition the set U , and Q will evolve to ensure the partition is determinate. But quantum measurement will partition U in such a way as to select precisely those topographies which are consistent with those observations and experiences, as long as such states are available — that is, as long as topographies which are consistent with those observations and experiences remain in N . For example, consider a neural topography which contains a population of cells whose only purpose is to force M to answer

“yes” whenever the question “Am I conscious?” is asked. The previous section showed that such a topography could not be consistent with intelligent processes which incorporate a capacity for honest deductive reasoning. But so long as such topographies remain within N , then quantum measurement will select precisely those topographies, and these topographies could be observed through neurological research. Indeed those topographies which *did* fully instantiate intelligent processes would be inconsistent with the conscious experience of self-certainty, and would therefore be selected out by the partition and hence *not included* in N . So the price one pays for consistency between conscious states and physical states is a lack of consistency between the selected topographies and the intelligent processes which supposedly supervene on them. One can see that in such a case the effect χ would not occur.

Of course in practice it is not just the particular sample of environmental situations which occur within a given human lifetime which one has to consider, but the sample of such situations which occur throughout the whole of human evolutionary history. As the range of actual environmental situations encountered by human beings throughout history becomes larger and larger, the permissible deviation of the topographies in N from perfect consistency with intelligent processes becomes smaller and smaller — just as, in the case of the oracle, the number of questions actually asked of the oracle gets ever larger, the oracle will have to get ever closer to true omniscience.

There are two potential difficulties with using single mind Q to “conceal” the effect χ . Firstly, mental operations such as deciding that one is conscious are not really like measurements of quantum observables. In the measurement of a quantum observable an eigenstate of that observable is selected randomly, in accordance with the quantum amplitudes associated with the various eigenstates. But in the specific example of deciding that one is conscious, only those neural states which are consistent with the outcome of that process are possible. Correlation of the observer’s physical state with the observer’s own mental state removes any possibility of quantum indeterminacy in this particular case.

Since clearly we must be correlated with our own brains this presents no problem for us. But consider an extraterrestrial visitor who is *not* correlated with our brain states or our mental states. Such a visitor would find it extremely peculiar that the usual rules of quantum indeterminacy were being flouted. One can see why by considering the example above of a population of cells whose sole purpose is to ensure that we always answer “yes” whenever the question “Am I conscious?” is asked. Such a neural topography, and the evolutionary history leading up to it, would be extremely unusual. An extraterrestrial visitor uncorrelated with our mental and brain states would expect to find many examples elsewhere in the universe of conscious beings whose brains did not exhibit such a topography. We would thus be unusual in being perhaps the only conscious beings in the universe who could be certain of being conscious, a circumstance which appears unreasonable.

One way round this problem would be to require that all intelligences in the universe, including all extraterrestrial intelligences, were in fact correlated with our own mental and brain states in some fundamental way. The source of such a correlation would presumably have to be found in the very early history of the universe. Another way would be to impose a requirement that the “minds” in single mind Q entail certain properties, and to require that the neural topographies they select be fully consistent with intelligent processes. In this second case the rules of quantum indeterminacy could be preserved, and χ would not be concealed and should be detectable.

The second problem is that single-mind Q in any case would not *completely* eliminate the possibility of χ . Consider a comprehensive program of neuroscientific research, as represented by a long sequence of measurements, completed before any attempt was made to detect χ . The result would be a subset V, which would be the intersection of all those subsets of U selected by their respective measurements. If the research program were intensive enough then V might be a very small subset indeed. In such a case, one could not be sure that V would still contain sufficient neural topographies, that at least one would remain which was consistent with the mental property of knowing that one is conscious. All neural topographies consistent with that outcome might have been partitioned out by the previous sequence of measurements. In such a case one would expect χ to be detectable subsequently.

Note that the subset V can be defined as follows:

$$V = U \setminus (V_n \cup V_e \cup W)$$

where V_n is the subset of U inconsistent with neuroscientific observations; V_e is the subset of U inconsistent with observed environmental properties; and W is the subset of U inconsistent with the existence of the non-physical “minds” required by single mind Q. The considerations in this section can be summarized by saying that, if the correct quantum statistics are to be maintained, then either all “minds” in the universe are correlated, or “minds” which are certain of their existence are found only on earth, or it is the case that the subset W is not empty.

Single Mind Q May Explain a Specific Operational Definition of Free Will

There is a sense in which the single mind Q approach to quantum mechanics may explain a certain notion of free will. To see how this is so, I now refer to some celebrated experiments by the physiologist Benjamin Libet. The first experiments of interest here refer to a phenomenon generally known as the *readiness potential* (Libet, 1985; Libet, Gleason, Wright, and Pearl, 1983). When human subjects are asked to time as accurately as possible when they experience the impulse to

perform a random movement, an EEG trace is observable up to 0.3 seconds before the subject's first conscious awareness of the impulse (this number is an average computed from aggregate data). This is known as the readiness potential. It might be argued that, since the EEG trace precedes the conscious impulse and in effect predicts it, the apparently random conscious impulse is not, in fact, random at all but determined by the neurophysiological state of the subject's brain. So, to the extent that one regards random impulses as a matter of free will, Libet's results can be taken as an argument against free will.

Libet's interpretation of these findings is controversial, particularly with respect to the readiness potential; and it is not my intention here to attempt to resolve this controversy. I wish to make the much narrower point, that even if the readiness potential can be regarded as a predictor of the subject's decision in a classical system, it cannot necessarily be regarded as such in a quantum system. The reason is that the neurological properties underlying the readiness potential may not actually have determinate values until the subject becomes consciously aware of her decision. In connection with this, an earlier experiment (Libet et al., 1972.) is of interest here. Using a technique known as *backward masking* which, for reasons of space, will not be described here, Libet found evidence that perceptual stimuli can take up to 0.5 seconds (with a minimum of 0.4 seconds) before they register as conscious impressions — it takes that long for the subject's brain to process them. This delay is called *perceptual latency*.

Single mind Q illustrates how the second effect may counteract the first. Consider an EEG machine which is in a superposition of two states: a state EEG_{ON} , in which the readiness potential is detected, and a state EEG_{OFF} in which no readiness potential is detected. These states are correlated with brain states $BRAIN_{ON}$ and $BRAIN_{OFF}$, in which the readiness potential occurs and does not occur respectively. From the perceptual latency effect described above, it will take roughly 0.5 seconds for the states EEG_{ON} and EEG_{OFF} to form a conscious impression in the mind of the observer reading the EEG machine — at which point, according to single mind Q, the superposition will be resolved to a single determinate state (albeit only in the minds of the conscious observers). But by that time, the subject's conscious awareness will already have selected a determinate value for the readiness potential, since the readiness potential is shorter than the perceptual latency.

In other words, it is impossible for any observer to perceive consciously if a readiness potential has in fact occurred, *before* the experimental subject experiences the conscious impression of a random impulse. Since in single mind Q determinate properties are mental properties, this means there simply *is no* determinate state for the readiness potential *or* the EEG trace before the subject becomes aware of her conscious decision. The readiness potential therefore cannot, *even in principle*, be used to predict the subject's decision before it happens. This will always be the case if the perceptual latency is longer than the

readiness potential. And so, according to single mind Q, it will be the subject who determines the state BRAIN_{ON} or $\text{BRAIN}_{\text{OFF}}$, and hence the state EEG_{ON} or EEG_{OFF} , by random selection. This state of affairs is empirically indistinguishable from the operational definition of free will posited by Libet, but removes any possibility that the readiness potential can be said to have a determinate value before the subject's conscious decision. Of course, this only applies to the rather limited sense of free will described by Libet. It is also subject to empirical review should subsequent research challenge the relative values of the readiness potential and perceptual latency.

What sort of neural mechanism might be implied by the effect described here? A neural network which exploits single mind Q might have the following properties: P is a population of cells, and I_1 and I_2 are, respectively, excitatory and inhibitory inputs to P. X is a population of cells I shall call the *state determiner* — population X determines the output of the network. E and Y are populations which are connected to X by reciprocal excitatory and reciprocal inhibitory connections respectively. X is connected to P by a delay line, which allows small changes in P to manifest before they are amplified by the connections from X to E and Y. $K(P)$ is the mean activity level⁵ of P, the value of which is equal to K_{idle} when $I_1 = I_2$. The network is set so that when the activity level of X is K_{idle} , both E and Y are inactive. An increase in the activity level $K(X)$ of X will drive $K(X)$ to a level K_{max} , and a decrease will drive $K(X)$ to K_{min} , which are respectively the maximum and minimum values of $K(X)$.

I now introduce a quantum noise term⁶ ϵ to P. (It is important to note that merely adding classical noise to the network will not work, since the effect being exploited here relies on the quantum superposition being maintained until a conscious decision is made.) I assume ϵ to be approximately Gaussian in distribution, with a mean of zero. Therefore when $I_1 = I_2$, the activity level of P will be:

$$K(P) = K_{\text{idle}} + \epsilon$$

The effect of this is to introduce a small variation in $K(P)$ which will quickly be amplified by the network so that the state determiner X will evolve to either K_{max} or K_{min} . In quantum mechanical terms, the state vector of the network

⁵Each cell in P, X, E and Y fires a number of action potentials within a certain time Δt . This number is assumed to follow a Poisson distribution with mean μ_k . Excitatory or inhibitory inputs are assumed to increase or decrease the value of μ_k .

⁶The most likely source of such noise is thought to be in the random variation of neurotransmitter release at neural synapses (Destexhe 2012). If these small random variations are considered equally likely to increase or decrease the likelihood that a cell will fire an action potential, then the cumulative effect of many such variations can be regarded as Gaussian-distributed with a mean of zero, if the number of effects is sufficiently large. It is unfortunately impossible to quantify these effects in any simple way since they depend critically on the internal connectivity of the network, and in particular on the extent of feedback connections within the populations of cells.

can be represented as a superposition of two states: a state MAX in which $K(P) = K_{\max}$, and a state MIN in which $K(P) = K_{\min}$. According to single mind Q, a single state, either MAX or MIN, will then be selected randomly once a conscious observation is made. (Different probabilities for K_{\max} and K_{\min} can be arranged by varying I_1 and I_2 so that $K(P)$ is initially either slightly greater or slightly less than K_{idle}).

Consciousness as a Fundamental Entity in Explanations of Nature

Finally I want to make a brief remark about how theories of consciousness, and the interaction of consciousness with the physical world, should include consciousness itself as an entity. Since consciousness cannot be fully decomposed into physical components, how can it be defined as a theoretical entity, and what properties should be attributed to it? The obvious starting point is to define consciousness in terms of precisely that property which turns out to be inconsistent with physical decomposition — that is to say, self-certainty. This property can be defined in terms of the mapping $E \rightarrow X$ which was set out in the first section of this article. If we refer to this mapping as the function p_0 , then self-certainty can be defined as the capacity of consciousness to perform the function p_0 with provable reliability and accuracy. This can be defined symbolically in terms of an infinite sequence of functions:

$$P_1, P_2, P_3, \dots$$

where every p_n can be defined in the following terms:

$$X_n = P_n(P_{n-1})$$

such that:

$$X_n = \text{YES if } p_{n-1} \text{ is performed accurately and reliably;}$$

$$X_n = \text{NO otherwise.}$$

Clearly, each function p_n in the sequence examines whether the previous function p_{n-1} has been correctly performed. These functions obviously correspond to the computations (or physical processes) described as part of the infinite regress in Argument A. However, unlike those processes, these functions are merely abstract representations of the *properties* of consciousness, and are not concrete entities in the physical world. In fact the representation of self-certainty in terms of a sequence of functions provides another way of proving the impossibility of self-certainty in a purely physical system, since it is easy to show that no physical system can perform all of these functions. To see why, one need only assume

some physical process P_n which performs each function p_n . If one assumes the functions are performed sequentially, then one notes that each P_n requires some time to execute, say δt . The infinite sequence of functions therefore requires a total time of δt multiplied by infinity. Alternatively if one assumes the various functions are performed in parallel, then each P_n requires some region of space, say δV , to execute. The total volume of space required to perform all the functions simultaneously is therefore δV multiplied by infinity. A physical system to perform the infinite sequence of functions would therefore need either to be infinitely large or to take an infinite amount of time, and neither contingency is physically reasonable.

The infinite sequence of functions can be summarized as a single function p_ω , identified by the subscript ω or *omega*:

$$X_\omega = p_\omega(E)$$

where:

$X_\omega = \text{YES}$ if it is provably the case both that $E = 1$ and p_ω is reliably performed;

$X_\omega = \text{NO}$ (or more accurately, is undefined) otherwise.

This is purely a notational convenience. One can regard a defining characteristic of consciousness as the ability to perform the function p_ω , and a defining physical property of consciousness as the χ effect, or violation of energy conservation, which is associated with it. Once defined, such a fundamental entity can be included in theoretical models, or simulations, of neurological or cognitive processes. This illustrates that it is not true, as is sometimes claimed, that allowing a non-physical basis for consciousness renders consciousness immune to analysis or understanding.

References

- Barrett, J.A. (1999). *The quantum mechanics of minds and worlds* Oxford: Oxford University Press.
- Caplain, G. (1995). Is consciousness a computational property? *Informatica*, 19, 615–619.
- Caplain, G. (2000). Is consciousness not a computational property? — Reply to Bojadziev. *Informatica*, 24, 79–81.
- Destexhe, A. (2012). *Neuronal noise*. New York: Springer.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–566.
- Libet, B., Alberts, W.W., Wright Jr, E.W., and Feinstein B (1972). Cortical and thalamic activation in conscious sensory experience. In G. Somjen (Ed.), *Neurophysiology studied in man* (pp. 157–168). Amsterdam: Excerpta Medica.
- Libet, B., Gleason, C.A., Wright, E.W., and 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*, 106, 623–642.
- Von Neumann, J. (1955). *Mathematical foundations of quantum mechanics* [R. Beyer, Trans.]. Princeton: Princeton University Press. (Originally published 1932)

On the Appearance and Reality of Mind

Demian Whiting

University of Hull

According to what I will call the “appearance-is-reality doctrine of mind,” conscious mental states are identical to how they subjectively appear or present themselves to us in our experience of them. The doctrine has had a number of supporters but to date has not received from its proponents the comprehensive and systematic treatment that might be expected. In this paper I outline the key features of the appearance-is-reality doctrine along with the case for thinking that doctrine to be true. I also defend the doctrine from some objections. Finally, I spell out the important metaphysical and epistemological implications of the appearance-is-reality doctrine of mind.

Keywords: appearance, reality, conscious mental state, phenomenology

When I am pained, I cannot say that the pain I feel is one thing, and that my feeling it is another thing. They are one and the same thing, and cannot be disjoined even in imagination. Pain, when it is not felt, has no existence.

Thomas Reid, 1855

Essays on the Intellectual Powers of Man

[T]he correspondence between a brain state and a mental state seems to have a certain obvious element of contingency. We have seen that identity is not a relation which can hold contingently between objects. Therefore, if the identity thesis were correct, the element of contingency would not lie, as in the case of heat and molecular motion, in the relation between the phenomenon (= heat = molecular motion) and the way it is felt or appears (sensation S), since in the case of mental phenomena there is no ‘appearance’ beyond the mental phenomena itself.

Saul Kripke, 1980

Naming and Necessity

It is sometimes held that there is no distinction between a conscious mental state and the way it is felt or appears. For instance, on this view the subjective appearance of pain — pain’s “painful” or “hurty” feel, in other words — *is* pain.

Many thanks to Paul Gilbert, Nick Zangwill, Raymond Russ, and three anonymous referees for their helpful comments on an earlier version of the paper. Correspondence concerning this article should be addressed to Demian Whiting, Ph.D., School of Politics, Philosophy and International Studies, University of Hull, Kingston-upon-Hull HU6 7RX, United Kingdom. Email: d.whiting@hull.ac.uk

I will call this the “appearance-is-reality doctrine of mind” or the “appearance-is-reality doctrine” for short. The appearance-is-reality doctrine is a striking thesis regarding the nature of conscious mental states, not least because there seem to be no other cases where the relationship between appearance and reality is one of identity. And as we will come to see, the implications of such a doctrine for metaphysics and epistemology are profound — so profound, in my view, that the appearance-is-reality doctrine is one of the most important theses in the philosophy of mind. So it is surprising that the doctrine has not received more explanation and defence from its supporters than it has to date. True the doctrine finds support with respect to at least some conscious states such as bodily sensations (see, for instance: Gertler, 2005; Horgan, 2012; James, 1890; Kripke, 1980; McGinn, 2004; Nagel, 1974; Reid, 1855; Searle, 1992, 1997). Moreover, it is a doctrine that is sometimes appealed to in defence of particular philosophical positions. Consider, for instance, Saul Kripke’s use of the doctrine when seeking to show that mind–brain identity theory is false. Nevertheless I am not aware of the doctrine, including its key features, having anywhere been spelt out in a systematic and comprehensive way. This is unfortunate because until this is done the doctrine risks being misunderstood and its philosophical implications are unlikely to be understood properly. Neither, in my view, has the reason for accepting the doctrine been explained very well. Is it a conceptual truism, so-to-speak, that pain, for instance, is its subjective appearance, or should we believe the doctrine for some other reason? Also the appearance-is-reality doctrine invites various objections that need answering (although as we will see, some of these rest on faulty understandings of the doctrine — another reason for getting clear on the details of the doctrine to begin with of course). And finally, it is my view that the philosophical implications of the appearance-is-reality doctrine have not been spelt out as well as they might be. The point of this paper, then, is to formulate and defend the appearance-is-reality doctrine, as well as outline its metaphysical and epistemological implications, and in the process show why that doctrine really is as important as some of us might think.

The Appearance-Is-Reality Doctrine of Mind

In its most concise form the appearance-is-reality doctrine of mind holds that all phenomenally conscious mental states are their appearances. Or, to put the point in another way, the doctrine holds that all phenomenally conscious mental states are the appearances of those mental states. I will now spell out four key features of the appearance-is-reality doctrine as just formulated.

The Doctrine Concerns Conscious Mental States

The appearance-is-reality doctrine concerns those mental states that we commonly think of as being phenomenally conscious — where a mental

state is phenomenally conscious if the mental state feels a certain way to us or if there is something that it is like to undergo that mental state. On my list I include sensory states such as pains, itches, tingles, emotions, and moods, but also more complex episodic intentional mental states, including conscious thoughts, desires, perceptual experiences, and imaginings.

It follows that the appearance-is-reality doctrine is not concerned with *non*-conscious mental states, if there are such things. I say “if there are such things” because, first, it is controversial whether some putative non-conscious mental states lack a phenomenology. For instance, not everyone thinks episodic thoughts are phenomenally conscious, whereas I along with a number of others take the view that thoughts are phenomenally conscious and, what is more, possess a distinctive phenomenology, one that is not to be described in terms of phenomenal properties associated with other mental states such as perceptual imagery (cp. Pitt, 2004; Strawson, 1994; see Bayne and Montague, 2011 for a summary of recent work on cognitive phenomenology) — and for that reason will speak of thoughts and thought-like states as being conscious.

Moreover, with respect to those states that are less controversially regarded as non-conscious, it seems to me unresolved as to whether such states are properly to be regarded as being mental by nature. For instance, should so-called dispositional states, such as dispositional beliefs, desires, and long-standing fears, be described as bona-fide mental states (cp. Strawson, 1994; see also Gertler, 2007)? The question is well-motivated because on the face of it, if to suffer a fear of dogs, for instance, is to be disposed to respond to dogs with episodic fear, then a fear of dogs does not look like a mental state as such, but only a disposition to undergo one, namely episodic fear (where episodic fear does have a phenomenology). Also and more intuitively, if we think there can be non-conscious mental states then we seem committed to the view that a mind can exist but for which everything is “dark” and that is an uncomfortable view to accept; for intuitively, a world in which consciousness does not exist is a world in which a mind does not exist, or, in other words, is a world in which that thing that makes the mental *mental* does not exist (cp. Kim, 1996, p. 237).

Still more will need to be said to satisfy everyone that all mental states are conscious and that is not the purpose of this paper. For that reason my position will be only that the appearance-is-reality doctrine concerns all conscious mental states, thus leaving for another time proper consideration of whether that doctrine ends up saying that *all* mental states are their appearances.

The Doctrine Concerns Phenomenal Appearances

By “appearance” (or equivalently “way of appearing”) I mean the *phenomenal* or *subjective appearance* of something, or, in other words, the way something feels or subjectively appears or manifests itself to mind — the *painful appearance*

or *feel* of pain, the *itchiness* of an itch, the *visual appearance* (or *look*) of a table,¹ and the *visual-like appearance* of an after-image, for instance. Intuitively and consistent with the appearance-is-reality doctrine, subjective appearances have the following two features. First, they have the property of being felt or appearing in a certain way. In what-it-is-likeness terms we might say there is *something that it is like* to have a painful or itchy feel, for instance — which is just to say these subjective feels or appearances are felt or have a certain feel to them. Second, phenomenal appearances have the property of being felt *in their entirety*. There is no part of a subjective appearance that does not feel a certain way or for which there is not something that it is like to undergo it. Again consider a painful or itchy feel. Notice how there is no part of the subjective appearance in question that does not have a certain feel to it. The painfulness of pain and the itchiness of an itch are felt in their entirety, so-to-speak.

There are a number of other things to be said here. First, the claim that subjective appearances are felt in their entirety does not entail on the appearance-is-reality doctrine that we have infallible knowledge of the nature of our conscious mental states. Although that claim has some positive epistemological implications (about which more later), it does not mean that we cannot be mistaken about how phenomenal appearances (and by implication the conscious mental states with which they are identical) feel in our experience of them. Pain's painful appearance might be apparent to us in all its phenomenal richness, but if, for instance, we are poor at describing the phenomenology, then we can still be led to form false judgments about pain's way of appearing (and, by implication, pain) — again I return to this point in the last section of the paper.

Second, it is important to distinguish phenomenal or subjective appearance from (what is commonly called) *epistemic seeming*, which is the way in which we might think or be inclined to think about something (see Chalmers, 1996, p. 190; Horgan, 2012; Schwitzgebel, 2008). Phenomenal appearances might provide evidential grounds for thinking or being disposed to think something is the case. For example, I might say that from the way the table looks I am inclined to think the table is rectangular in shape or that from the painful feel of pain

¹In fact, with respect to mind-independent objects, there are two meanings of the term “appearance” (or “look” or “taste” and so on) that it might be useful to distinguish here. First, there is the meaning intended in the text, where to talk about an object's appearance is to talk about the way an object feels or subjectively appears in our experience of it. This is the phenomenal sense of “appearance.” But second, there is a sense of “appearance” where to talk about the appearance of an object is to talk about the way the object would feel or subjectively appear in our experience of it if the object were perceived. This latter sense of “appearance” is what we have in mind when we say that an object has an appearance even though it is not being perceived by anyone (consider how we might speak of the “look” of the watch hidden in my drawer). But note it is with respect to mind-independent objects only that “appearance” can have these two meanings. Assuming the truth of the doctrine, the reason for this is straightforward. If pain is the way it subjectively feels or appears to us, then it makes no sense to speak about the way pain would feel if it were experienced, as that would imply that pain can exist when it is not being experienced (which it cannot if pain is its subjective appearance).

I am inclined to judge that pain is a raw feeling (on this point, see Horgan, 2012, p. 407). But phenomenal appearances are not epistemic seemings. Indeed, the examples just given are saying something informative only because phenomenal appearance and epistemic seeming have different meanings. It follows that the appearance-is-reality doctrine is not the thesis that for a person to undergo a conscious mental state is for that person to think or be disposed to think he or she is undergoing a conscious mental state. The reason why this distinction is important will become evident later when I address certain objections to the doctrine.

Third, it is important to recognise that to talk about the phenomenal appearance of something is not the same as talking about the way something appears to us to be, or, what I take to mean the same thing, the properties something appears to us to have (on this point, see Thompson, 2009). For instance, to speak about the subjective feel or appearance of a table — the table looking rectangular from here, say — is not the same as talking about the properties a table appears to me to have. One way of coming to see the difference between phenomenal appearance and the way something appears to be is by considering what each would entail if objects commonly thought to be non-mental admitted of no appearance/reality distinction. Now, if that were the case and “appearance” were read as “phenomenal appearance” then it would follow that all of reality is *mental* reality. If we can say of worldly objects — say concrete tables and chairs — that to be is to appear in the phenomenal sense of “appear” (that is to say, if we can say of worldly objects that to be is for them to be *felt*), then the result is idealism of one variety or other. But the same is not true if “appearance” were taken to mean “the way something appears to be.” For then, although it would follow that we have certain insight into the nature of the world (since we are supposing the world just as it appears to us to be), it would not follow that all of reality is mental reality; for it might be the case that worldly objects appear to us to be mind-independent — indeed, I believe I speak accurately in saying that worldly objects do appear to most people to be mind-independent — and, therefore, *are* mind-independent if they just are as they appear to us to be.

Another way of coming to see the distinction between phenomenal appearance and the way something appears to be is simply by noting that the claim that pain appears to be the way that pain feels, promises to tell us something about the nature of pain or the sort of thing pain is, namely that pain is its subjective feel or way of appearing. But that claim could not promise to tell us any such thing if there were no distinction between the two senses of “appearance” in question; for then that claim would be saying that pain appears to be the way pain appears to be, which although trivially true would in itself tell us nothing about what pain might actually be.

It follows that the appearance-is-reality doctrine is not the thesis that conscious mental states are the way they appear to be. The point is important because it

means that the appearance-is-reality doctrine is to be seen as a metaphysical thesis (that is, a thesis about what conscious mental states are, namely their subjective appearances) and not an epistemological thesis regarding the status of our knowledge of a mental state (which that thesis would seem to be if were taken to be saying that conscious mental states are the way they appear to us to be). Moreover, once we recognise the distinction just drawn, we will be able to understand much better the implications the appearance-is-reality doctrine has for mind-brain identity theory; for it is as a metaphysical thesis that the appearance-is-reality doctrine causes serious difficulties for that theory.

The Doctrine Concerns the Phenomenal Appearances of the Mental States in Question

To say that conscious mental states are their appearances or ways of appearing is to say the subjective appearances with which conscious mental states are identical are none other than their own appearances or ways of appearing. Now of course this is consistent with holding that different conscious mental states manifest themselves subjectively in different ways. And that there is variation in how different conscious mental states feel or subjectively appear to us is made evident when we reflect on the phenomenology of different conscious mental states. For instance, we come to see that pains, visual experiences, thoughts, imaginings, and emotions differ from one another in terms of how they feel. For example, whereas some conscious mental states seem to phenomenally manifest a non-intentional or non-object-directed nature (pains, itches, and moods perhaps), other conscious mental states (thoughts, imaginings, and perceptual experiences, for instance) manifest an object-directed nature.

Moreover, corresponding to differences in the feel or subjective appearances of different conscious mental types, we find differences in the feel or appearances of token instances of the same mental types. For instance, my thought that Paris is the capital of France shares with my thought that Smith is a salesman, an object-directed feel or appearance. However, the object-directed appearances of those thoughts differ in terms of how the world is presented in mind; whereas the first thought presents *Paris as being the capital of France*, the second presents *Smith as being a salesman*.

It is not always easy to describe how different conscious mental states feel or subjectively appear to us. Some conscious mental states have appearances that are so fine-grained or elusive that they fail to admit of easy description (consider the phenomenology involved in the visual experience of a complex landscape). But this is not to say these mental states do not admit of differences in subjective appearance (and indeed we are often able to know this even in cases where we find it difficult to say what characterises a given appearance or distinguishes it from other appearances). Again the implications that this has for the epistemology of mind are explored more fully below.

Does the claim that conscious mental states are the subjective appearances of those mental states threaten regress? On a superficial reading of that claim it might be thought to threaten regress because if conscious mental states are their appearances, then the appearances with which conscious mental states are identical must be identical to *their* appearances, and so on ad infinitum. But the worry of regress could be well-founded only if it were supposed that conscious mental states were somehow distinct from the way they subjectively appear to us, as then it would be the case that appearances will proliferate. But of course, the appearance-is-reality doctrine denies this. Since, according to that doctrine, the relationship between the conscious mental state and its way of appearing is one of strict identity (as I spell out in more detail, below) the threat of regress does not arise.

The Doctrine Holds that Conscious Mental States Are Nothing More than Their Phenomenal Appearances

To say that something *is* its subjective appearance is to say that that thing is *nothing more than* its subjective appearance or that it is *exhausted by* its subjective appearance. So, for instance, to say that pain is its painful feel or appearance is to say that pain is nothing more than its painful feel or way of appearing or that pain is exhausted by its painful feel or way of appearing.² According to the appearance-is-reality doctrine the relationship between conscious mental states and their subjective appearances is one of strict identity.

This is a striking feature of the doctrine. In no other case is the relationship between object and appearance normally considered to be one of identity. Tables and chairs, for instance, are not normally considered to be identical to the way they look or smell or feel in our experience of them. In their case, the relationship between the appearance and reality seems to be a non-constitutive, probably causal one. Thus we might say that tables and chairs are causally responsible for the way they feel or subjectively appear to us but they are not composed of or identical to their appearances. But the appearance-is-reality doctrine holds the same is not true in the case of phenomenally conscious mental states and their subjective appearances. That thesis holds that in the case of a conscious mental state the appearance really *is* the reality.

²I recognise not everyone thinks pain is its painful feel. For instance, it might be held that when we introspect pain we find that pain comprises (also) certain emotions such as anger and displeasure, or that pain has a motivational nature — so perhaps pain is the imperative "behave differently!" (see, Klein, 2015). Personally I think any account that does not identify pain with its painful feel fails to be faithful to the phenomenology of pain, and for this reason I will continue to talk about pain in the way I do. But note, that as it stands, the disagreement is solely one regarding the nature of pain's appearance — as to whether, for instance, pain presents itself as a painful feeling or as an emotion or as an imperative to behave some way (or perhaps as a compound of all three). It is not a disagreement regarding the claim that pain is its subjective appearance (whatever the nature of that appearance might be), which is the claim that principally concerns us.

It follows the appearance-is-reality doctrine is at odds with any theory of phenomenal appearance that identifies the way a mental state feels with that mental state's external relational properties. Consider Michael Tye's theory of phenomenal content, which holds that a mental state's phenomenal properties are its representational properties, where by "representational properties" Tye means a mental state's causal-convariational properties (see, for instance, Tye, 2000, 2005). Tye thinks that pain's feel or phenomenal character is just a matter of pain representing some bodily condition, where that is a matter of pain standing in the right causal relation with the bodily condition that it represents. According to Tye it is because pain causally correlates with bodily damage (or more precisely, causally correlates with bodily damage in optimal conditions) that pain represents bodily damage, and it is the representing of bodily damage that *is* pain's phenomenal character.

But this is at odds with the appearance-is-reality doctrine because if the way pain feels is a causal relation holding between pain and some bodily condition, then the way pain feels can be no part of pain itself. Indeed, on a causal/functional story pain itself might turn out to be nothing more than a physiological state (a firing of C-fibres say) albeit one standing in the right causal relations with other physiological activity (for instance, states of bodily damage) and/or bodily behaviour. If the appearance-is-reality doctrine is true then the appearance of a conscious mental state cannot be a causal/functional property because the appearance could not then *be* that conscious mental state.

Can this be made consistent with the claim that a mental state's intentional properties are themselves phenomenal properties? The claim that they are phenomenal properties was made above where it was held that the possession of an intentional or object-directed nature seems constitutive of the phenomenology of many conscious mental states, for instance, my thought that Paris is the capital of France. I think the right thing to say here is that if by a mental state's intentional properties we mean "representational" in the causal/functional sense of "representation," then on the appearance-is-reality doctrine a mental state's representational properties could not be part of its appearance (as this would prevent us from identifying mental states with their appearances). But I along with other philosophers (for instance, Chalmers, 2003, 2004a; Horgan and Tienson, 2002), take the view that this is not the only way of understanding "representation." Although there is a non-phenomenal sense of "represent" that picks out a mental state's causal/functional properties, there is another *phenomenal* sense — "phenomenal intentionality," as it is sometimes called — that picks out *the way in which the mind presents the world to itself*. This latter sense of "represent" picks out not a causal-functional property but rather a property that is phenomenally manifest in episodic thoughts and some other conscious mental states and thus a property that *is* constitutive of some ways of appearing, namely those ways of appearing that possess an intentional or object-directed nature.

The Case for the Appearance-Is-Reality Doctrine

What is the argument for the appearance-is-reality doctrine? Is it a conceptual truism that conscious mental states are indistinguishable from the way they feel? Certainly many people's concept or idea of a conscious mental state is consistent with the appearance-is-reality doctrine. For instance, it does seem to be part of many people's idea of pain that pain has a painful feel and is indistinguishable from the way it feels. This provides some support for the doctrine. In particular, any account of a conscious mental state that does too much violence to commonly held intuitions regarding a conscious mental state is likely to raise serious doubts as to whether it is still a conscious mental state about which we are speaking.

Nevertheless, there is reason not to rely solely on people's concepts or intuitions when seeking to understand what something is. These can be incomplete or mistaken with respect to how they represent something. Even if many people's concept or idea of pain is that pain is its painful feel, it is an open question as to whether the referent of the term "pain" is the way many people conceive pain to be (and not, for instance, the way someone who does not share that view conceives pain to be).

A more compelling argument, then, for the appearance-is-reality doctrine and one that vindicates commonly held intuitions, appeals directly to our experience or observation of a mental state. That is to say, I think we are justified in claiming that conscious mental states are the way they feel because that is evident from our experience of those mental states. One way of coming to see this is by reflecting on what remains when a pain ceases to be painful as might happen if an analgesic is taken. I think it is evident from our experience or observation of the pain in question that we are left with *nothing at all*, in much the same way that it is evident from our experience or observation of some physical object that if we remove certain of its physical properties — its spatial properties, for instance — the object itself ceases to be.³ Another way is by reflecting on what experience tells us when there is a change in how a conscious mental episode feels, when, for instance, an episode of pain feels very painful to begin with but less so as time goes on. Again I think it is evident from observing a mental state that when there is a change in how a mental state *feels*, the nature of the *mental state* itself changes. In much the same way it is evident from our observation of some physical object that if we alter certain of its physical properties the nature of the object changes also.

³This is why we are justified in thinking there are no such things as unfelt pains (or unfelt feelings) — for again, observation of pain makes evident for us the fact that pain when it is not felt simply ceases to be or is no pain at all.

And here it is useful to contrast the case of conscious mental states with other items in the world. Thus it is in no way evident from our observation of tables and chairs or water and heat, for instance, that what we are observing are the ways these things feel to us in our experience of them. Indeed, I think it is evident from our experience of other items in the world that they are *not* their ways of appearing; for it is evident from our observations of such items that they are part of a mind-independent world. Thus we come to see that these items are self-subsistent entities, so-to-speak, entities that can endure when not felt by us and which do not undergo change in virtue of change to the way they feel to us in our experience of them.

Now, to be clear, the argument is not we are justified in thinking conscious mental states are the way they feel because that is evident from the way they feel. This is no argument of course because we cannot tell on the basis of how a mental state feels whether the mental state *is* the way it feels. Rather, the argument is that it is on the basis of our *experience or observation of a conscious mental state* that we come to see that the mental state and its way of appearing are one and the same. Or equivalently, the argument is that it is on the basis of our *experience of the way a conscious mental state feels* — as opposed to the way a mental state feels — that we come to see that the mental state and its way of appearing are one and the same.

We need to distinguish, then, between “an experience of a conscious mental state” (or “an experience of the way a conscious mental state feels”) and “the way a conscious mental state feels.” An experience or observation of a mental state comprises an *introspective representation* of a mental state, whereas the way a mental state feels is a property of the mental state that is part of an experience of the mental state (hence the locution: the way a mental state feels in our experience of it) but is not a representation of the mental state.

The appearance-is-reality doctrine holds that conscious mental states are the way they feel in our experience of them, not that mental states are our experiences of them. This is a good thing too; for as I explained above it is evident from the subjective appearance of some conscious mental states that those mental states lack intentional objects, whereas those mental states could not be without objects if they *are* mental representations. However, it is the observation or experience of a conscious mental state and not the way a conscious mental state feels in our experience of it that justifies thinking a conscious mental state *is* the way it feels; for again, when we experience or introspectively observe a conscious mental state we are able to see by means of our encounter with or direct cognitive access to the mental state in question that the thing we are observing (or “looking at,” so-to-speak) is nothing other than its own manner or way of appearing.

The argument, then, is not that conscious mental states are the way they feel because that is evident from the way they feel; for our experience of a mental state is not the same as the way it feels. Nevertheless, we might wish to inquire further

into what it is that justifies thinking the experience of a conscious mental state delivers to us truths about that mental state; for even if the experience of a mental state is not the same as how the mental state feels, what is to say the experience does not tell us truths only about how the mental state feels? Now, if it were the case that the experience tells us truths only about the way conscious mental states feel then by the same token we ought to say our experience of other items in the world tells us truths only about the way those items feel — and that is clearly mistaken. For instance, my experience of water tells me truths about *water* — including the fact of water being a watery substance — and not truths about water's way of appearing. What then justifies thinking an experience of a mental state tells us truths about the mental state? The answer must be that it is the mental state that is being represented by us when we experience it. Thus our experience of pain gives us information about pain — including the fact of pain being its own way of appearing — because it is pain we are observing when we experience pain; hence it is pain we glean certain truths about.

And if it is asked what justifies thinking that it is *pain* we are observing and not pain's way of appearing? Of course if the appearance-is-reality doctrine is true, the answer is: nothing justifies thinking this. Pain is its own way of appearing; therefore, to experience pain is to experience pain's way of appearing. But suppose we take the question to be probing the assumption that what we are observing is pain and not pain's way of appearing *only*? In fact it does not matter much whether what we are observing is pain; for the argument seeks to show only that whatever it is we are observing that thing is its own way of appearing. Suppose the thing we are observing is pain's way of appearing but not pain. In that case the argument would be that it is evident from our observation of the thing that is pain's way of appearing that that thing is its own way of appearing. This would give us the appearance-is-reality doctrine with respect to the subjective appearance — where the reality is the subjective appearance and the appearance is the subjective appearance's way of appearing. [And it is only a small step from there to say that really it is pain's way of appearing that is the conscious mental state, not pain itself.] Nevertheless I do not think the appearance-is-reality doctrine is true for pain's way of appearing but not for pain, since I think we are wholly right to think it is pain we are observing, not pain's way of appearing only. The question is: How can we be confident that what we are observing is *pain*? And the answer has to be: because the thing we are observing is what picks out the referent of the term "pain." And we know this because we have been taught to use the term "pain" to refer to that thing which we find on experiencing or observing it to be its own manner of appearing — in much the same way that we know what we are observing is *water* because we have been taught to use the term "water" to refer to that which we find when observing it to be a watery substance.

Objections and Replies

In what follows I address two types of objection to the appearance-is-reality doctrine of mind. First, there are those objections that provide positive reasons for denying that conscious mental states are their subjective appearances. These objections proceed by describing counter-examples to that doctrine or by seeking to show the doctrine is false for conceptual reasons. The second type of objection does not seek to show that the doctrine is false, but questions our confidence in the grounds for accepting the doctrine in the first place, the implication being that if there is reason to doubt those grounds, then even if the doctrine is true we may still lack justification for believing it to be true.

To begin with, it might be held that cases can be described where the putative identity between a mental state and the subjective appearance of that mental state does not exist. First, there are those cases in which there is the subjective appearance but no conscious mental state. For instance, Rosenthal (2005) observes that some dental patients report themselves to be in pain (owing to such things as anxiety and the non-painful sensation of vibration) but where physiological factors make it clear that no pain can be present (we might imagine patients have been anaesthetised, for instance; see also Brown, 2010; Churchland, 1988).

One response to such cases is disbelief: if it seems to us that we are in pain then we must be in pain! I have some sympathy with this response. In the normal case we should take people's sincere reports to be in pain at face value. But there is another reply available to us, which can allow that in exceptional cases we can be mistaken about whether we are in pain (say owing to unusual cognitive pressures) but which does not mean rejecting the appearance-is-reality doctrine. This is because it can be argued that such counter-examples equivocate between epistemic seeming and subjective appearance, a distinction that was made earlier in the paper. In other words, when we imagine such cases we imagine people who, despite thinking themselves to be in pain, do not experience anything pain-like, and, therefore, if the appearance-is-reality doctrine is true, are *not* in pain, just as Rosenthal holds.

Other counter-examples argue to the opposite view, namely that there can be a conscious mental state but no appearance. For instance, some have the intuition that one can have a headache all day but only be intermittently aware of it. But in reply, it is not clear what to say about such cases. With respect to the all day headache example, if the claim is that one can have a headache without *thinking one has a headache* then that is not a problem for the appearance-is-reality doctrine because that doctrine is not saying that to have a headache is to think one is having a headache. As we have seen, that would confuse the epistemic use of "appear" with the phenomenal use

of “appear.” It would also have other unattractive implications; for instance, it would seem to mean that young children cannot have headaches since they lack the conceptual abilities needed to think they are having a headache.

On the other hand, if by an unfelt headache we mean a headache that lacks the characteristic headache phenomenology, then the intuition there can be pain without subjective appearance looks question-begging. Of course, we sometimes speak of having a headache all day, which might seem to suggest that our heads can ache when there is no pain phenomenology, but it is difficult to interpret such speech to be saying anything other than that our heads were aching off and on all day (compare, people sometimes say they have not stopped eating when what they usually mean is they have eaten at regular intervals, not that they have eaten *non-stop!*) Barring further argument, counter-examples such as the all day headache one, end up begging the relevant questions (for useful discussion of the all day headache example, see Robinson, 2004).

A final counter-example makes appeal to the so-called “transparency thesis” that many philosophers of perception accept (for a summary and discussion of the relevant literature, see Kind, 2003). For that thesis might be taken to threaten the claim that conscious mental states are their appearances, since the transparency thesis might seem to suggest that what is apparent to us when we have a perceptual experience is *that which we perceive* and not the experience itself. Now if this is correct then perceptual experience is a counter-example to the appearance-is-reality doctrine. This is because it would follow that perceptual experiences cannot be how they subjectively appear to us, since such experiences do not appear to us in any way whatsoever (for again it would be the objects and their properties that appear to us, not the experiences themselves).

But I think we should reject the claim that perceptual experiences do not subjectively appear to us. Of course it is the object and its properties that are apparent to us *in* our experience of the object, so-to-speak. For instance, when I perceive a red ball, it is the ball and its redness — and not my experience of the ball and its redness — that are apparent to me in the experience of the red ball. However that is not to say that when I perceive a red ball, the experience of a red ball is not apparent to me. For there is a distinction to be drawn between the object of perception, on the one hand, and the perceptual experience itself, on the other — and although the latter is not apparent in my perceptual experience of the object (hence it is not the object being visually perceived) it might appear to me all the same. And indeed I think that when I visually perceive a red ball the experience of the red ball is apparent to me. For my visual experience of the red ball comprises the *red ball visually appearing or looking some way* to me (on this point, see Siewert, 2004), and the red ball looking some way to me is no less evident to me when visually experiencing the red ball than the red ball itself. But if that is the case then we do not have a counter-example to the doctrine;

for it follows that our perceptual experiences are apparent to us in addition to the things being represented in those experiences.⁴

The view that conscious mental states are the way they feel faces a conceptual objection. The objection is that this view has to be false because to speak about the way something feels is to speak of a *relation* — call it the “feeling-a-certain-way relation” — that holds between the mental state and the subject of the mental state in question. But it might be claimed that no relation can be identical with one or more of its relata. In reply I agree that to speak of the way something feels is to speak of a relation but think this would disprove the appearance-is-reality doctrine only if the relation in question were an external one, that is to say, a relation that is external to or not identical with one or more of the relata. In the case of mind-independent objects and the way they feel, it is clear the relation is an external one. For instance, the coin I am visually experiencing is independent of its looking elliptical to me — where a justification for saying this, is that it is possible for the coin to enter into a different relation (for it to look different from the way it does look to me) from the one it does in fact enter.

However, in the case of mental states and their ways of appearing, the relation in question seems to be an *internal* not external relation. This is because conscious mental states do not seem to be able to exist independently of or prior to the way they feel to us. For instance, the pain I feel in my hand does not exist independently of or prior to the way my pain feels to me — where a justification for saying this is that it is not possible for my pain to enter into a different relation (that is, for my pain not to feel the way it does feel to me) from the one that my pain does in fact enter. The conceptual objection would succeed only if there were reason to hold that all relations must be external relations. But although many relations are external it is not clear why all relations must be external. Indeed, the appearance-is-reality doctrine is one counter-example to the thesis that all relations are external; for that doctrine holds plausibly that in the case of a conscious mental state the feeling relation *is* constitutive of one of the relata, namely the conscious mental state that feels a certain way to us. Therefore, unless a compelling case can be given for holding that all relations are external, the conceptual objection also ends up begging the question.

⁴One way of building on this point is by considering what Amy Kind calls “exotic cases” such as blurry vision and after-images (see Kind, 2008 for a detailed discussion). Suppose I have poor vision but am looking at a red ball whilst wearing glasses. Due to the power of the lenses my visual experience of the red ball is crystal clear. Now suppose I remove the glasses resulting in the red ball looking blurry to me. The red ball looking blurry is something that is now apparent to me, and this differs from what was apparent to me when I was wearing glasses. Is what is apparent to me here the thing that I am visually experiencing? The answer seems to be negative; for the thing that I am visually experiencing — namely the red ball — does not seem to change after I remove my glasses, whereas what is apparent to me here — namely the red ball looking blurry to me — differs from what was apparent to me before removing my glasses. So what is it that is apparent to me? Again the answer seems to be: my experience of the red ball, or, the red ball looking some way to me.

A second type of objection allows that for all we know, conscious mental states might be their subjective appearances, but complains that the grounds for accepting that doctrine — namely that that doctrine is vindicated by our experience or observation of conscious mental states — are unstable or inconclusive. Although this type of objection does not seek to show the doctrine is false, it does try to demonstrate that agnosticism regarding the truth of the doctrine is the only justified position to adopt (at least until a better argument for thinking the doctrine is true can be given).

To begin with, one might doubt the reliability of forming beliefs about the nature of mental states on the basis of our experience of them. Such a doubt might be motivated by consideration of the following cases. First, there are those cases where, owing to unusual cognitive pressures, people are led into forming erroneous beliefs on the basis of their experience of a mental state. Consider the dental case described already, where due to anxiety and the non-painful sensation of vibration people are led to believe erroneously that they are in pain (Rosenthal, 2005; see also Churchland, 1988). But although such cases pose a challenge to the view that we are infallible regarding our knowledge of our mental states (and it is the infallibility thesis to which such cases are normally cited as an objection), it is difficult to see how such cases show that we are not justified in claiming that conscious mental states are the way they feel on the basis of our experience of them. After all, when we judge from our experience of pain that pain is the way it feels, we will normally be making that judgement when there are no unusual cognitive pressures that could risk leading us into error.

Other reasons for doubt appeal to less out of the ordinary cases. For instance, some commonly had mental states are very detailed in their presentation and others are short-lived or have rapidly changing natures; such features can create difficulties when reflecting on our experience of a mental state for the purpose of gaining a deeper understanding of that mental state (see Schwitzgebel, 2008). I suggest later in the paper that with respect to such mental states we can often attain good insights into what they are like. But even in cases where reflecting on our experience of a mental state delivers us limited knowledge about that mental state, such cases seem to pose little threat here. This is because, first, when we say it is evident from an experience of a mental state that the mental state in question is the way it feels, we will normally be basing what we say on the experience of a mental state the phenomenology of which is relatively easy to grasp — the painful feel of pain, for instance. But second, even with respect to mental states that are more complex in terms of how they feel, it seems to me that we are normally able to see that the mental state in question *is* the way it feels, even if we cannot say fully *how* it feels. I might not be able to describe very well how a pang of nostalgia feels, for instance, but that does not stop me from being able to see from my experience or observation of an episode of nostalgia that that mental episode is its way of appearing.

Another objection also questions the grounds for accepting the appearance-is-reality doctrine but not because of doubts regarding our ability to make accurate judgments about what the experience of a conscious mental state might be telling us, but rather because of concerns regarding the possible limitations of the experience or observation itself with regards to what it is able to tell us about our conscious mental states. There are two forms this objection might take.

First, it might be held that even if our experience of a mental state can tell us about *that which we can experience*, still it might not be able to tell us everything about a mental state, including, for instance, the mental state's neural-physiological properties. This seems to be what Patricia Churchland has in mind when she says that “not *everything* about the nature of pain is revealed in introspection — its neural substrate, for example, is not so revealed” (Churchland, 1998, p. 117; italics in original). Now, in reply, if by “neural substrate” Churchland means something on which a mental state *depends* then even if our experience of a mental state can tell us nothing about *that*, this would be no objection to the view that it is our experience of a mental state that justifies us holding that mental states are their ways of appearing; for on that view there is no reason to suppose our experience of a mental state is able to tell us anything about that which is not part of a mental state, including anything about that on which a mental state might depend. But if as seems more likely by “neural substrate” Churchland means something that constitutes a mental state, then how we reply will depend on the nature of the constitution-relation in question. Thus if the view is that pain, for instance, might be composed in part of how it feels and in part of a neural state — a neural-phenomenal compound, so-to-speak — then it is difficult to avoid the worry that anyone who holds that view fails to understand accurately the referent of the term “pain.” It would be similar to thinking that water comprises more than a watery substance, a watery substance plus something else. To believe that would be to fail to understand that “water” picks out that which we can see from our experience or observation of it to be a watery substance only. Similarly if someone held that pain comprises more than the way it feels — the way it feels plus a neural state — then we would be right to complain that the individual fails to understand that “pain” refers only to that which we can see from our experience or observation of it to be pain's way of appearing.⁵

On the other hand, if the view is that a conscious mental state is indeed nothing more than the way it feels, but qua the way it feels that mental state is a neural state, then although I think the appearance-is-reality doctrine rules that idea out (as I will argue in the next section), it suffices to point out that that view does not

⁵It is worth remarking that such a view still gives us the appearance-is-reality doctrine with respect to the conscious parts of conscious mental states. That would also be a striking doctrine and would have much the same metaphysical and epistemological implications as the ones I outline later in the paper (which makes the motivation for endorsing this view, as opposed to the view that conscious mental states in their entirety are the way they feel, even more puzzling).

undermine the appearance-is-reality doctrine nor the grounds for thinking that doctrine is true. This is because even if it were the case that mental states qua their subjective appearances are neural states (as a mind–brain identity theorist might hold), it would still be the case that conscious mental states are nothing more than the way they feel, and it would still be the case that we are justified in thinking this because the truth of that claim is evident from our experience or introspective observation of our mental states.

The other form of the objection under consideration holds that for all we know the experience is unable to tell us anything about our mental states; for it might be claimed that it is possible the nature of a mental state is wholly unavailable to experience — in much the same way, for instance, it might be held the atomic or molecular nature of water is not evident to us when we observe water. But, in response, to say the nature of water is not evident to us when we experience water is not the same as saying that our observation of water tells us nothing about water. Although the experience is unable to tell us about water's atomic structure it is able to tell us certain important truths about water. For instance, it is able to tell us that water is a watery substance. And likewise I have argued the experience is able to tell us certain important truths about our conscious mental states, including the fact that our mental states are their ways of appearing.

Implications of the Appearance-Is-Reality Doctrine of Mind

Metaphysical Implications

It might be held that the appearance-is-reality doctrine need not threaten mind–brain identity theory because even if pain, for instance, is identical to its painful feel, this is consistent with holding that pain qua the way it feels is a neural or physiological state, a firing of C-fibres, say. But I want to now show why mind–brain identity theory is not sustainable if we accept the appearance-is-reality doctrine along with a plausible thesis regarding the nature of neural–physiological states.

The argument for holding that conscious mental states are not identical to neural or physiological states or activity is as follows.

1. Every phenomenally conscious mental state is the subjective appearance of that state
2. No neural state is the subjective appearance of that state

Conclusion: No phenomenally conscious mental state is a neural state

The first premise is a statement of the appearance-is-reality doctrine. I explained that the primary justification for thinking that conscious mental states are indistinguishable from their subjective appearances is that this is evident to us from our experience or observation of them. In support of premise 2, the view might

be advanced that neural states do not subjectively appear to us in any way; for if that view is true then it cannot be the case that neural states *are* how they appear to us.⁶ But for the sake of argument, suppose that the neural states in question do subjectively appear to us in some way. For instance, let us suppose that the feeling associated with pain captures how the firing of C-fibres subjectively appears to us (in much the same way that it might be held that a feeling of heat, for instance, captures the way in which heat feels or appears to us).

Nevertheless, a strong argument can be given for thinking that the second premise is still true, which is as follows. If a neural state is its subjective appearance — for instance, if C-fibre firing is the way it feels when we are in pain — then it cannot have a nature that is not apparent to us in the experience of that state. This is because to describe the way something feels is to describe a property that is apparent to us in the experience of the thing. Therefore, if something is the way it feels or subjectively appears to us — *exhausted by* its appearance, *nothing more than* the way it feels or subjectively appears to us — then the nature of that thing must be apparent to us in our experience of it. But in the case of neural states, we find they possess a microphysical nature that is not apparent to us in the experience of those states. Although we can theorize about the microphysical properties or structure of a neural state, that structure is not evident to us in our experience of that state (in much the same way that the microphysical properties of heat and water are not evident to us in our experience of heat and water). But then it follows that it cannot be the case that a neural state is its subjective appearance. For that to be the case its nature would need to be apparent to us.

And it will not help to respond that neural states might be identical to their subjective appearances because the subjective appearances with which neural states are identical might have a nature that is not evident to us. If something *is* a property that it possesses and which is apparent to us — for instance, if a neural state is the way it feels in our experience of that state — then the nature of the property with which the thing is identical will be apparent or evident to us as well (for the thing and its property are one and the same, and, therefore, their natures, the properties composing them, must be the same also). Neither will it help the mind-brain identity theorist to reply that the microphysical properties of a neural state are apparent to us in our experience of a neural state albeit under

⁶In fact, it would be sufficient to demonstrate the truth of premise 2 to show only that neural states are not necessarily felt (see Kripke, 1980, pp. 152–154; Nagel, 1974), a view I sympathise with, since I sympathise with the stronger claim (namely that neural states are not felt at all) which entails the weaker one (namely that neural states are not necessarily felt). However, neither the stronger nor weaker claim on its own is likely to worry much the mind-brain identity theorists, since they are likely to argue that both claims express nothing more than certain modal intuitions, which might be in error. The argument I go on to give in the main text then seeks to show the truth of premise 2 in a way that does not appeal to such intuitions (although that argument is clearly consistent with, and, if successful, possibly explains and justifies commonly-held intuitions regarding the relationship between mind and brain).

a phenomenal mode of presentation. If the microphysical properties of a neural state are apparent to us, then it should be possible to know those properties just by attending to our experience of the neural state; otherwise it would not be the case that those properties are apparent (or evident) to us. But again, that information is not evident to us in our experience of a neural state (or any other physical state, for that matter) — for instance, the number of atoms that make up a neural state is not evident from our experience of a neural state — and for that reason the present reply will not succeed either.

The conclusion is an application of Leibniz's law. If conscious mental states but not neural states are comprised of how they feel or subjectively appear, then it follows that conscious mental states have different properties from neural states and, therefore, are not neural states. And the argument goes through even assuming a functionalist view of mind. This is because most functionalists claim that mental states are neural states that possess the right causal/functional properties. But that is not to say that mental states are not neural states; rather it is to say that the neural states with which mental states are identical qualify as mental states only if they possess the right functional properties. Thus most functionalists are committed to a mind–brain identity theory (for instance, the view that pain is C-fibres firing) and, therefore, open to the objection just outlined.⁷ We can state the objection to functionalism as follows:

1. Every phenomenally conscious mental state is the subjective appearance of that state
 2. No functional state (= neural state with causal/functional properties) is the subjective appearance of that state
- Conclusion: No functional state is a phenomenally conscious mental state

As the quotation at the beginning of the paper shows, Kripke draws also on the appearance-is-reality doctrine when arguing against mind–brain identity theory. However, it is clear that what is supposed to be doing most of the work in Kripke's argument is the modal intuition that the relation between mind and brain is contingent, hence not one of identity. The reason Kripke appeals to the lack of an appearance/reality distinction in the case of a mental state and its appearance

⁷Of course, functionalism does not require the relevant states to be physical (cp. Levin, 2009). What matters is the functional role played by those states and not the nature of the states themselves. Thus for the functionalist a non-physical state could qualify as a mental state if it has the right causal/functional properties. Now, there would be something odd about a type of functionalism that sought to identify the relevant states with something non-physical, as that would remove the motivation for adopting a functionalist view in the first place (which I take to be its pretension to describe mental states in a purely physical way). But setting that issue aside, it suffices to say that the sort of functionalism with which I am concerned is the sort that does consider the functional states in question to be physical, as I am concerned only to show why the appearance-is-reality doctrine entails that mind–brain identity theory is false.

is to ward-off an objection, which tries to explain away the modal intuition by appealing to the idea that that intuition is motivated by a failure to distinguish the appearance from the mental state itself — in much the same way the intuition that the relation between water and its microphysical properties is contingent is (arguably) based also on a failure to distinguish water from its appearance (the way it looks or feels). But if the relation between appearance and reality is one of identity, as Kripke thinks, then the element of contingency cannot be eliminated in this way, and, therefore, the inference from the modal intuition to the claim that mind-brain identity theory is false is not blocked.

My argument against mind-brain identity theory differs from Kripke's argument since I do not appeal to modal intuitions. Modal intuitions are not mentioned in either premise and the considerations offered in support of those premises make no reference to modal intuitions. Indeed, it seems to me that both premises are true for a posteriori reasons. So on my view, once we accept the appearance-is-reality doctrine (on the grounds that its truth is evident to us from our experience of conscious mental states), we need only make the plausible (again empirically motivated) opposite claim about the nature of neural activity (namely that neural activity is *not* its appearance) to show that mind-brain identity theory is false. Now, it is not possible here to say whether the argument I give is more successful than Kripke's argument or other arguments against mind-brain identity theory that rely on modal intuitions (see, for instance, Chalmers, 1996, 2003). Nevertheless, it might be pointed out that because my argument does not rely on modal intuitions then it cannot be objected that that argument is unstable because it derives substantive metaphysical conclusions from epistemological premises (see, for instance, Loar, 1990; Papineau, 2002), and in that respect at least, the argument I give seems to be in a dialectically stronger position than arguments that appeal to modal intuitions.

Epistemological Implications

In what follows, I spell out two epistemological implications of the appearance-is-reality doctrine. To begin with, let us recap something that was said above, namely that although the appearance-is-reality doctrine entails that conscious mental states are felt in their entirety — and that, therefore, their natures are laid bare for us — this is consistent with holding that we might describe them incorrectly when reflecting on them, and as a result have false beliefs regarding those mental states. The appearance-is-reality doctrine, then, does not entail that we have infallible knowledge of our mental states, and in that respect our epistemic position with respect to mental states seems no better-off than our epistemic position with respect to other things. And yet, that conscious mental states are their subjective appearances has at least one positive epistemological implication. Although that feature of a mental state does not entail that we

might never be mistaken about the mental state that feature does entail that if we describe the appearance correctly then necessarily we describe the reality correctly (for again the appearance is the reality). And in this respect our epistemic position with respect to mental states seems a lot better-off than our epistemic position with respect to other things; for unless we are idealists about the physical world we will hold that physical objects are distinct from the way they feel and that although the appearances of physical objects gives us grounds for thinking objects have certain properties (for instance, a table looking rectangular to us in our experience of the table gives us reason to think the table is rectangular), those appearances do not legitimize the logical inferences we are entitled to make in the case of conscious mental states and the way they feel (thus it might turn out the table is not rectangular).

In other words then, if the appearance-is-reality doctrine is true, whereas we can say that we are able to know with certainty that if something has a painful way of appearing (or, in other words, feels the way pain feels to us) then that thing *is* pain or that if something has a thought-like way of appearing (or, in other words, feels the way a thought feels to us) then that thing *is* a thought, we cannot say that we know with complete certainty that if something visually appears to us as water visually appears to us, then that thing must be water, or that if something feels the way heat feels to us then it must be heat (as the examples of “twin-water” and “twin-heat” attest). In the case of water and heat (and all other non-conscious phenomena) there exists forever an ontological gap between the object and the appearance and along with that there exists forever the possibility that the subjective appearance might not correspond to or be a true measure of the reality.

The second epistemological implication is that the lack of an appearance/reality distinction entails that attention to the phenomenology will be inescapable if we wish to understand the nature of conscious mental states. Our concepts of a mental state seem inadequate to that task because these can be mistaken or incomplete regarding how they represent the mental state. And neither, if the argument I spelt-out against mind-brain identity theory is successful, can empirical methods of inquiry that investigate the brain and its neural activity tell us much if anything about the nature of mental states (though they might tell us important things about the physical states or properties on which mental states depend in some way). Phenomenology as a method of enquiry into the nature of a mental state is inescapable because it is only by reflecting on how a mental state feels in our experience of the mental state that we are able to examine the nature of the mental state itself (and by so doing be in a position to vindicate or develop or reject outright a concept or belief about that mental state).

This gives rise to the question of whether the indispensability of phenomenology as a method of inquiry into the nature of our mental states is something that should concern us. Certainly it is likely some philosophers will think it should (see, for instance, Dennett, 1991; Schwitzgebel, 2008, 2011). This is because there

are a number of obstacles to describing the phenomenology, including but not limited to: difficulties in isolating mental states or attending to them properly (owing to such things as their short-lived and changeable natures); the concern that people's descriptions are vulnerable to bias or conceptual incompetence (Horgan, 2012); and the concern that some mental states might be ineffable or indescribable (see Chalmers, 2004b; Schwitzgebel, 2008, 2011).

These concerns succeed in showing that we need to proceed with care when seeking to understand our conscious mental states and to be mindful of the obstacles that can exist. And, crucially, we need to be aware of the fact that our descriptions (and especially those of our more complex or elusive or not so commonly had mental states) can be inadequate or incomplete and in need of revision. But phenomenology is a justified form of inquiry for all that (as well as being the only form of inquiry into the nature of mental states that we have available to us). Although there are a number of things that can interfere with our ability to describe the appearance of conscious mental states, it is difficult to see how these need always or even typically undermine this ability. To begin with, many conscious mental states are so familiar and pervasive in everyday waking-life or so obvious in their presentation that it seems unjustified to state that our descriptions of them are unreliable. Does anyone really think we have reason to doubt that pain manifests a feeling quality, or that a greenish visual-appearance presents differently from a reddish visual-appearance, or that an episode of anxiety feels different from an episode of sadness?

Furthermore, even in cases where it is difficult to describe the way a mental state subjectively appears there are a number of tools available to help facilitate sound phenomenological insights. For instance, if we are having difficulty in describing the subjective appearance owing to theoretical bias or an inability to isolate a mental state, then discussion with others can go some way to overcome such obstacles. X says an episodic fear of an object presents as a feeling that is directed at an object. Y questions this, stating she thinks an episodic fear of an object manifests as a non-intentional feeling *along with* a thought directed at an object. Y is raising the possibility that X has failed to notice that the mental state in question is compound, comprising a non-intentional feeling *and* a thought. This disagreement can prompt both to revisit the subjective appearance and evaluate which of the two descriptions is more faithful to the phenomenology. I wish not to adjudicate between X and Y here (but for my way of treating such mental states, see Whiting, 2011, 2012), but indicate only how discussion with others can help us to attain better phenomenological insights. Or if the reason why we cannot describe the phenomenology is owing to our limited powers of description, then there are various linguistic tools and methods (such as the use of metaphor) that can go some way to overcome that problem, not to mention non-linguistic forms of representation, such as painting and music. Of course it might be the case that philosophers are not always the most skilful at providing the words or

forms of pictorial representation that best describe the phenomenology of our more ineffable human experiences, but then I take it no one thought the task of describing complex human experience is one for philosophy alone anyway.

References

- Bayne, T., and Montague, M. (2011). Cognitive phenomenology: An introduction. In T. Bayne and M. Montague (Eds.), *Cognitive phenomenology* (pp. 1–34). Oxford: Oxford University Press.
- Brown, R. (2010). Deprioritizing the a priori arguments against physicalism. *Journal of Consciousness Studies*, 17, 47–69.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers, D. (2003). Consciousness and its place in nature. In S. Stich and F. Warfield (Eds.), *Blackwell guide to philosophy of mind* (pp. 102–142). Oxford: Blackwell.
- Chalmers, D. (2004a). The representational character of experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford: Oxford University Press.
- Chalmers, D. (2004b). How can we construct a science of consciousness. In M. Gazzaniga (Ed.), *The cognitive neurosciences III* (pp. 1111–1119). Cambridge, Massachusetts: MIT Press.
- Churchland, P. M. (1988). *Matter and consciousness*. Cambridge, Massachusetts: MIT Press.
- Churchland, P. S. (1998). Brainishy: Nonneural theories of conscious experience. In S. Hameroff, A. Kaszniak, and A. Scott (Eds.), *Towards a science of consciousness II: The second Tuscan discussions and debates* (pp. 109–126). Cambridge, Massachusetts: MIT Press.
- Dennett, D. (1991). *Consciousness explained*. London: Penguin.
- Gertler, B. (2005). In defence of mind–body dualism. In J. Feinberg and R. Shafer–Landau (Eds.), *Reason and responsibility: Readings in some basic problems of philosophy* (pp. 285–296). Belmont: Thomson Wadsworth.
- Gertler, B. (2007). Overextending the mind? In B. Gertler and L. Shapiro (Eds.), *Arguing about the mind* (pp. 192–206). New York: Routledge.
- Horgan, T. (2012). Introspection about phenomenal consciousness: Running the gamut from infallibility to impotence. In D. Smithies and D. Stoljar (Eds.), *Introspection and consciousness* (pp. 403–422). Oxford: Oxford University Press.
- Horgan, T., and Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 520–533). Oxford: Oxford University Press.
- James, W. (1890). *The principles of psychology, Volume 1*. New York: Henry Holt and Co.
- Kim, J. (1996). *Philosophy of mind*. Oxford: Westview Press.
- Kind, A. (2003). What’s so transparent about transparency? *Philosophical Studies*, 115, 225–244.
- Kind, A. (2008). How to believe in qualia. In E. Wright (Ed.), *The case for qualia* (pp. 285–298). Cambridge, Massachusetts: MIT Press.
- Klein, C. (2015). *What the body commands. The imperative theory of pain*. Cambridge, Massachusetts: MIT Press.
- Kripke, S. (1980). *Naming and necessity*. Cambridge, Massachusetts: Harvard University Press.
- Levin, J. (2009). Functionalism. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/entries/functionalism>. Accessed on 14 March 2016.
- Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108.
- McGinn, C. (2004). *Consciousness and its objects*. Oxford: Oxford University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Clarendon Press.
- Pitt, D. (2004). The phenomenology of cognition or *what is it like to think that p?* *Philosophy and Phenomenological Research*, 69, 1–36.
- Reid, T. (1855). *Essays on the intellectual powers of man* (sixth edition). [J. Walker, Ed.]. Boston: Philips Sampson and Company.
- Robinson, W. (2004). *Understanding phenomenal consciousness*. Cambridge: Cambridge University Press.
- Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

- Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117, 255–273.
- Schwitzgebel, E. (2011). *Perplexities of consciousness*. Cambridge, Massachusetts: MIT Press.
- Searle, J. (1992). *The rediscovery of mind*. Cambridge, Massachusetts: MIT Press.
- Searle, J. (1997). *The mystery of consciousness*. London: Granta Books.
- Siewert, C. (2004). Is experience transparent? *Philosophical Studies*, 117, 15–41.
- Strawson, G. (1994). *Mental reality*. Cambridge, Massachusetts: MIT Press.
- Thompson, B. (2009). Senses for senses. *Australasian Journal of Philosophy*, 87, 91–117.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, Massachusetts: MIT Press.
- Tye, M. (2005). Another look at representationalism. In M. Aydede (Ed.), *Pain: New essays on its nature and the methodology of its study* (pp. 99–120). Cambridge, Massachusetts: MIT Press.
- Whiting, D. (2011). The feeling theory of emotion and the object-directed emotions. *European Journal of Philosophy*, 19, 281–301.
- Whiting, D. (2012). Are emotions perceptual experiences of value? *Ratio*, 25, 93–107.

A NOTE ON OUR BOOK REVIEW POLICY

We will accept book reviews for publication each issue. Authors wishing to submit book reviews are urged to write with the above interdisciplinary framework firmly in mind. All books *solicited* from publishers will be sent to selected individuals for review. JMB also accepts unsolicited reviews. Reviews should be absent of all titles except the name of the work reviewed, author of work reviewed, place of publication, publisher, date of latest publication, number of pages, and price. Any individual wishing to submit a review should contact our Book Review Editor for further information: Steven E. Connelly, Ph.D., Department of English, Indiana State University, Terre Haute, Indiana 47809. Email: steven.connelly@indstate.edu

JMB is abstracted or indexed in: *Cultures, Langues, Textes: La Revue de Sommaires*; *Current Contents (Social and Behavioral Sciences)*; *EBSCOhost*; *EMBASE/Excerpta Medica*; *International Bibliography of Book Reviews*; *International Bibliography of Periodical Literature*; *Linguistics and Language Behavior Abstracts*; *Physics Abstracts*; *Psychiatric Rehabilitation Journal*; *PsychINFO/Psychological Abstracts*; *Research Alert*; *Social Science Citation Index*; *Social Work Abstracts*; *Sociological Abstracts*; *The Philosopher's Index*. The Journal of Mind and Behavior website is located at www.umaine.edu/jmb/.

The Journal of Mind and Behavior

Winter 2016

Vol. 37 No. 1

CONTENTS

- Which Identification is Disturbed in Misidentification Syndromes? A Structural Analysis of Fregoli and Capgras Syndromes** 1
Stéphane Thibierge and Catherine Morin
- Does Functionalism Offer an Adequate Account of Cognitive Psychology?** 15
James M. Stedman, Thomas L. Spalding,
and Christina L. Gagné
- Consciousness is Not a Physically Provable Property** 31
Catherine M. Reason
- On the Appearance and Reality of Mind**..... 47
Demian Whiting