STICH

AND HIS CRITICS

Edited by

Dominic Murphy
and Michael Bishop
STICH

AND HIS CRITICS
Philosophy is an interactive enterprise. Much of it is carried out in dialogue as theories and ideas are presented and subsequently refined in the crucible of close scrutiny. The purpose of this series is to reconstruct this vital interplay among thinkers. Each book consists of a temporary assessment of an important living philosopher’s work. A collection of essays written by an interdisciplinary group of critics addressing the substantial theses of the philosopher’s corpus opens each volume. In the last section, the philosopher responds to his or her critics, clarifies crucial points of the discussion, or updates his or her doctrines.

1. Dretske and His Critics  
   *Edited by Brian McLaughlin*

2. John Searle and His Critics  
   *Edited by Ernest Lepore and Robert van Gulick*

3. Meaning in Mind: Fodor and His Critics  
   *Edited by Barry Loewer and Georges Rey*

4. Dennett and His Critics  
   *Edited by Bo Dahlbom*

5. Danto and His Critics  
   *Edited by Mark Rollins*

6. Perspectives on Quine  
   *Edited by Robert B. Barrett and Roger F. Gibson*

7. The Churchlands and Their Critics  
   *Edited by Robert N. McCauley*

8. Singer and His Critics  
   *Edited by Dale Jamieson*

9. Rorty and His Critics  
   *Edited by Robert B. Brandom*

10. Chomsky and His Critics  
    *Edited by Louise M. Antony and Norbert Hornstein*

11. Dworkin and His Critics  
    *Edited by Justine Burley*

12. McDowell and His Critics  
    *Edited by Cynthia Macdonald and Graham Macdonald*

13. Ernest Sosa and His Critics  
    *Edited by John Greco*

14. Stich and His Critics  
    *Edited by Dominic Murphy and Michael Bishop*
STICH

AND HIS CRITICS

Edited by

Dominic Murphy
and Michael Bishop
Contents

Notes on Contributors vii

Introduction 1
    DOMINIC MURPHY AND MICHAEL BISHOP

1 Is There a Role for Representational Content in Scientific Psychology? 14
    FRANCES EGAN

2 Representationalism Reconsidered 30
    PETER GODFREY-SMITH

3 On Determining What There Isn’t 46
    MICHAEL DEVITT

4 Eliminativism and the Theory of Reference 62
    FRANK JACKSON

5 Why Isn’t Stich an Eliminativist? 74
    FIONA COWIE

6 A Defense of the Use of Intuitions in Philosophy 101
    ERNEST SOSA

7 Reflections on Cognitive and Epistemic Diversity: Can a Stich in Time Save Quine? 113
    MICHAEL BISHOP

8 Simulation Theory and Cognitive Neuroscience 137
    ALVIN GOLDMAN
CONTENTS

9 The Triumph of a Reasonable Man: Stich, Mindreading, and Nativism 152
   KIM STERELNY

10 Against Moral Nativism 167
   JESSE J. PRINZ

11 Replies 190
   STEPHEN STICH

List of Publications by Stephen Stich 253

Index 265
Notes on Contributors

Editors

Dominic Murphy is Senior Lecturer in History and Philosophy of Science at The University of Sydney. He works in philosophy of psychology, biology, and neuroscience. He is the author of *Psychiatry in the Scientific Image* (2006).

Michael Bishop is Professor of Philosophy at Florida State University. His principal research areas are epistemology and philosophy of science. He is the author, with J. D. Trout, of *Epistemology and the Psychology of Human Judgment* (2005).

Contributors

Fiona Cowie is Associate Professor of Philosophy at The California Institute of Technology. She works at the interface of philosophy and the sciences of the mind, and she is the author of *What's Within* (1998).

Michael Devitt is Distinguished Professor of Philosophy at The Graduate Center, City University of New York. He has published many books and articles in the philosophy of language, epistemology, and metaphysics. His most recent book is *Ignorance of Language* (2006).

Frances Egan is Associate Professor of Philosophy at Rutgers, The State University of New Jersey. Her recent work concentrates on the role of representational content in computational models of cognitive capacities. She also has research interests in the theory of vision and in computational neuroscience.

Peter Godfrey-Smith is Professor of Philosophy at Harvard University. He works mainly in the philosophy of biology and the philosophy of mind. He is the author of *Complexity and the Function of Mind in Nature* (1996).
Alvin Goldman is Board of Governors Professor of Philosophy and Cognitive Science at Rutgers, The State University of New Jersey. His principal research areas are epistemology, philosophy of mind, and cognitive science. His most recent book is *Simulating Minds* (2006).

Frank Jackson is Distinguished Professor of Philosophy at The Australian National University. He also holds appointments at Princeton and La Trobe Universities. He is the author of *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (1998). He was awarded the Order of Australia in 2006.

Jesse J. Prinz is Distinguished Professor of Philosophy at The Graduate Center, City University of New York. He has research interests in cognitive science, philosophy of psychology, philosophy of language, moral psychology, and esthetics. His is the author of *The Emotional Construction of Morals* (2007).

Ernest Sosa is Board of Governors Professor of Philosophy at Rutgers, The State University of New Jersey. He has published in many areas of epistemology. His most recent book is *A Virtue Epistemology* (2007).

Kim Sterelny is Professor of Philosophy at Victoria University in Wellington, and in the Philosophy Program at The Research School of Social Sciences, The Australian National University, Canberra. He is the author of *Thought in a Hostile World: The Evolution of Human Cognition* (Blackwell, 2003) and, with James Maclaurin, of *What is Biodiversity?* (2008).

Stephen Stich is Board of Governors Professor of Philosophy and Cognitive Science at Rutgers University and Honorary Professor of Philosophy at the University of Sheffield. He has written extensively on issues in cognitive science, philosophy of language, philosophy of mind, epistemology, moral theory, and philosophical methodology. In 2007, he received the Jean Nicod Prize sponsored by the French Centre National de la Recherche Scientifique, and in 2008 he was the first recipient of the American Philosophical Association’s Gittler Award for outstanding scholarly contribution to the philosophy of the social sciences.
Since the late 1960s, Stephen Stich has produced a substantial body of philosophical work that is remarkable for its range and power. With clarity and verve, Stich has defended positions and arguments that are bracing and often contrary to conventional philosophical wisdom. He has tackled an astonishingly diverse set of philosophical topics – most notably, linguistics and innateness, mental content and eliminativism, pragmatism and epistemology, moral psychology, genetic engineering, the nature of “mindreading” (how we ascribe mental states to ourselves and others), experimental philosophy, and issues of philosophical method. Two further aspects of Stich’s body of work are noteworthy. His writing sets a standard of lucidity and clarity of exposition that is too rarely met in philosophy today. What’s more, philosophy for Stich is a genuinely collaborative activity. His first publication (1968) was jointly authored, and since then he has collaborated with almost 40 different people – and counting!

The Stich oeuvre might appear to be a menagerie of disparate topics lacking any clear guiding philosophical theory or organizing principle. But perhaps the best way to frame Stich’s work over the years is in terms of the assumption that philosophical problems are best understood as arising naturally out of the sciences. If we want to know about the nature of mind, morality, language or culture, we should pay close attention to what the sciences tell us about mind, morality, language or culture. Of course, Stich’s research testifies to the idea that solutions to philosophical problems cannot simply be “read off” the science. Stich (sometimes together with his collaborators) has offered a number of well-known arguments for thinking that certain specific argument-types commonly employed by philosophers are unlikely to be fruitful (Weinberg, Nichols and Stich 2001; Mallon, Machery, Nichols and Stich, forthcoming). But given the bold naturalism implicit in Stich’s philosophy, some might be surprised to discover that Stich has never offered a positive, general account of his philosophical method. This should not really be surprising, however. Why would we expect a naturalist to engage in a priori reflections about the nature of philosophy? (Indeed, it is ironic how much commentary on philosophical naturalism comes from deep in the anti-naturalistic armchair.) In Stich’s writings, we see a philosopher tackling important problems using an extraordinarily diverse range of considerations and techniques culled from philosophy and the sciences. This kind of
methodological opportunism is precisely what we should expect from a naturalist who understands that in philosophy, as in science, we can have good reasons for thinking that certain methods are ineffective but we can’t always know what methods will be successful until after we try them out. The clear-headed naturalist understands the futility of trying to identify a priori the correct philosophical methods with any high degree of precision.

The articles in this volume represent most of the major themes Stich has addressed over the course of his career. So perhaps the best way to introduce those themes is to give a brief synopsis of these articles and Stich’s replies to them.

In *From Folk Psychology to Cognitive Science (FP)*, Stich explored the relationship between our commonsense, “folk” accounts of human behavior, which appeal liberally to representational states like beliefs and desires, and the accounts of human behavior to be found in cognitive science. He argued that the “manifest image” and the “scientific image” of the human mind clash rather dramatically. Stich went on to argue for eliminativism – the view that these representational states do not exist. A central feature of Stich’s arguments for eliminativism was his Protagorean account of how we ascribe content to folk psychological states: When I ascribe to you the belief that snow is white, I am saying that you are in a belief state that is similar to the belief state that would normally underwrite my assertion that snow is white (1983: 88). Stich argued that this view of content ascription implies that the representational states of folk psychology are dogged by various forms of vagueness, context-sensitivity, and observer-relativity. And as a result, the folk psychological concept of belief “ought not to play any significant role in a science aimed at explaining human cognition and behavior . . . [and] despite appearances to the contrary, [it] does not play a role in the best and most sophisticated theories put forward by contemporary cognitive scientists” (1983: 5). Both Frances Egan and Peter Godfrey-Smith aim to explore the issue of where things stand today – what role do representational states play in contemporary cognitive science?

Egan begins her article by noting that while Stich argued for eliminativism in *FP* in 1983, sixteen years later, in *Deconstructing the Mind (DM)* (1999), Stich is more guarded. In *DM*, Stich is prepared to concede that intentional categories might have a useful scientific role to play. Egan thinks that they do, and defends her claim by showing how intentional categories play important roles in two broad areas of cognitive science. The first set of examples derives from computational cognitive psychology. While such theories individuate cognitive systems in terms of their abstract computational roles, it is only by appealing to semantic, intentional categories that we can explain how those roles hook up with the environment. For example, while we can individuate a cognitive system purely in terms of its intrinsic mathematical properties, the only way to account for the fact that it is the *visual* system is by characterizing it as having certain intentional properties. The very context-sensitivity of intentional ascription, says Egan, is what allows us to understand a system in terms of its environment, since two systems that share a computational characterization might have evolved in different environments to do different jobs. While computational psychology must appeal to semantic states, it does not make use of folk psychological states like beliefs and desires. Egan argues that these folk psychological states do play a role in other areas of psychology, notably developmental and social psychology.
The basic move Stich makes in reply to Egan is to note that philosophers have distinguished a number of different notions of representation (Cummins 1989; Ramsey 2007). In FP, Stich argued that the notion of representation underlying folk psychology would play no role in a mature cognitive science – where he took that notion of representation to be vague, context-sensitive, and observer-relative. In replying to Egan, Stich argues that the notion of representation underlying computational cognitive psychology is not this folk notion. It has none of the idiosyncrasies of the folk notion – instead, it is “determined by the distal properties tracked by [the mechanism’s] internal states” (p. 22). With respect to Egan’s second set of examples, Stich grants that the folk psychological notions of representation are appealed to in some thriving areas of social psychology (although Stich has some misgivings about Egan’s particular examples). But he is wary of supposing that these areas of psychology presuppose a single, folk psychological notion of representation, much less the notion he described in FP. Some of the sources of Stich’s wariness become apparent in his response to Peter Godfrey-Smith.

Godfrey-Smith begins by noting that “cognitive science makes free use of concepts of representation, knowledge, reasoning, and so on” (p. 31). And yet philosophers have not succeeded in showing how an organism’s representational states can be explained in terms of its physical or biological states. Godfrey-Smith suggests that the reason for this is that the cognitive sciences may advert to systems that are representational in a wide variety of different ways: “A physical system can acquire a well-motivated semantic description by a variety of ‘paths’” (p. 32). A central kind of representation is the basic representational model (BRM), a notion derived from model-based conceptions of understanding discussed in recent philosophy of science. Godfrey-Smith discusses a paradigm example of BRM, one in which a “reader mechanism” takes one thing, M, to be a guide to another thing, W. The representation M (say, a street map) is used by a reader mechanism (the map reader) to be a guide to W (say, the location of the bus station). To be guided, the reader mechanism will often exploit some kind of resemblance between M and W – but the BRM allows for many different kinds of resemblance. “All the quirkiness, vagueness, and context-sensitivity of the notion of ‘resemblance’ are supposed to be in play here.” Models can be “unclear, indefinite, and changeable” about how and to what degree they resemble the target (p. 33).

Stich largely agrees with the general thrust of Godfrey-Smith’s views about representation, although he reserves judgment on many particulars. For example, Stich’s Protagorean account of how we ascribe content to folk psychological states implies that those states are riddled with vagueness, context-sensitivity, and observer-relativity. So Godfrey-Smith can be seen as largely confirming Stich’s views about the theoretical disorderliness of representation. Stich grants, however, that Godfrey-Smith might be right that Stich’s arguments for eliminativism were perhaps too pessimistic about the ability of cognitive scientists to productively employ the messy and unwieldy explanatory category of representation.

Michael Devitt and Frank Jackson consider a line of argument that appears in Deconstructing the Mind (DM 1996). There, Stich casts a skeptical eye on the central argument he had proposed for eliminativism:
Beliefs (and propositional attitudes generally) are the central posits of folk psychology.

Folk psychology is a seriously mistaken theory.

Therefore, beliefs (and propositional attitudes generally) do not exist.

This argument requires an extra premise, namely, that the central posits of seriously mistaken theories do not exist. In defending eliminativism, Stich had assumed this tacit premise was warranted on the basis of a description theory of reference, according to which a theoretical term (like ‘belief’) refers to whatever is picked out by the descriptions the theory associates with that term. If, as Stich argued, folk psychology is seriously mistaken, then the descriptions associated with its central posits do not pick out anything. And so those central posits do not exist. William Lycan (1988) responded by rejecting the description theory of reference in favor of a causal theory of reference. Such theories hold that reference is some sort of causal relation between the uses of a term (like ‘belief’) and some thing or type of thing in the world. A distinctive feature of causal theories is that they allow for the possibility that theoretical terms can successfully refer even though they are embedded in deeply mistaken theories (e.g., early theories of the atom). So if we assume a causal theory of reference, then even if Stich was right that folk psychology is seriously mistaken, it is still possible for its central terms, like ‘belief’, to successfully pick out something in the world.

Lycan’s argument woke Stich from his “dogmatic slumbers.” In DM, Stich does not attack Lycan’s argument head-on by defending a description theory of reference. Rather, Stich notes that a debate about ontology (Are there beliefs?) has turned into a debate about language (How do theoretical terms refer?). Stich argues that it is a mistake to turn issues about ontology into issues about reference. “What exactly are the facts that a theory of reference is supposed to capture?” Stich asks (DM: 6). He offers two possible answers. First, it might aim to capture some “folk theory” of reference – our commonsense views about the link how words pick out things in the world. Or second, the “proto-science” project holds that the goal of a theory of reference is “to characterize a word–world mapping that will be useful in one or another empirical discipline such as linguistics, cognitive psychology, or the history of science” (DM: 6). Stich thinks neither project is likely to succeed. The folk theory project isn’t appropriate because there is no reason to believe that our “folk theory” of reference, in all its details, is true. And the proto-science project is suspect because it is unlikely that any area of science really employs a word–world reference-like relation in its explanations and, even if scientific explanations do appeal to such relations, there is no reason to think that every area of science will employ the same unique word–world relation (DM: 44–5).

A natural objection to this line of argument is this: How are we to resolve difficult ontological issues except by appealing to theories of reference? Suppose those things we call ‘cats’ are really robots and not mammals (Putnam 1970, 1975). What would be the right thing to say – that cats exist but they aren’t what we thought they were or that cats don’t exist? Stich suggests a number of ways such hard cases might be handled without benefit of a theory of reference: some hard cases might simply remain unresolved because no one is interested in whether the posit exists; other cases might be resolved via implicit previous agreements about what conditions would have to obtain in order for a posit to
exist; still others might be decided by social and political factors internal to the relevant science, or by broader social and political factors (DM: 65–71).

Michael Devitt agrees with much of this. He agrees that it is a mistake to resolve ontological problems by appeal to theories about language. Philosophers “should adopt the slogan, ‘put metaphysics first’” (p. 48). Devitt also agrees that the folk theory project is a non-starter because there is no reason to think the folk are right about the nature of reference. But Devitt argues that Stich’s arguments concerning the proto-science project are too “gloomy.” Devitt distinguishes two components of the proto-science project. The first involves characterizing the reference relation. There is no reason to be gloomy about this task. Our folk intuitions permit us to (fallibly) identify uncontroversial examples of reference and then we construct a theory about “what is common and peculiar to” reference (p. 49). The resulting theory can lead us to revise our original intuitions. The second component of the proto-science project is showing how reference is scientifically useful. Once again, Devitt argues that Stich’s discussion of this issue is too gloomy. The science of linguistics includes semantics and “prima facie, semantics does have a role for reference because the most popular theories of meaning are truth-referential.” And so “the meaning of the sentence ‘Bush is deceitful’ is largely explained by the sentence’s being true iff there is something that ‘Bush’ designates and ‘deceitful’ applies to. So, the meaning of the sentence is partly explained in terms of two referential relations, designation and application” (p. 53).

Stich carefully teases apart the various issues on which he and Devitt agree and disagree – expanding and revising his original argument in subtle ways. The main point of contention concerns the prospects for constructing a theory that identifies “what is common and peculiar to” reference. Stich raises a number of objections. For example, he asks us to consider the Twin Earth and Godel/Schmitt intuitions that are standard in the reference literature (Putnam 1975; Kripke 1972). Either such intuitions will be included in the store of uncontroversial evidence that is to be used to decide between theories of reference or they will not. If they are not included, then the evidence will be too feeble to decide between competing theories. If they are included, Stich argues that these identifying judgments are not “uncontroversial” because people in different cultures seem to have systematically different reference intuitions about these crucial cases (Machery et al. 2004).

Unlike Devitt, Frank Jackson agrees with Stich that the proto-science project is hopeless. Jackson argues that “the best policy is to be easy-going about what counts as representation” in the sciences (p. 65). The representation relations of interest to evolutionary biologists are bound to be different from those that are of interest to weather forecasters. To privilege one of these relations as the reference relation smacks of “dogma.” But Jackson does think that “there had better be” a special reference relation that is relevant to ontological questions – that is relevant to whether atoms, species, aether, or beliefs exist (p. 67). The crucial step in Jackson’s solution to Stich’s challenge is here:

Our thoughts have (reasonably) determinate contents which we can report using sentences. Anyone who denies this cannot allow that they know what this paper, or indeed any paper . . . is about . . . We use sentences to pass on what we think and to find out what
others think, and we often do so successfully. It follows that the word–world relations which
we have just noted are properly counted as cases of reference are very often determinate (or
determinate enough) and known. (p. 70)

The reference relation that is relevant to any ontological claim (e.g., there are electrons)
is the one that captures the content of the relevant belief (e.g., the belief that there are
electrons). Of course, this leaves open the possibility of purely verbal “no fault” disagree-
ments – disagreements about, for example, whether some plant is a shrub or a tree
(p. 72). But the important point, for Jackson, is that this account of reference legitimately
grounds all manner of non-trivial ontological arguments, including Stich’s argument for
the nonexistence of propositional attitudes.

The original problem arose for Stich because philosophers seem to resolve ontological
issues (Are there beliefs?) by appeal to semantic issues (What does “belief” refer to?). Stich’s
challenge is that philosophers who formulate theories of reference seem to be embarked
on one of two projects, and on neither project does it make sense to say that a theory of
reference can help to resolve ontological disputes. According to Stich, Jackson’s solution
to this problem merely moves the bump in the rug to another place. “To oversimplify a
bit, on Jackson’s account, the reference of terms in natural language is determined by the
reference of thoughts.” But “if one accepts Jackson’s general view of the relation between
language and thought, then there is no need to say or write anything at all in posing the
eliminativists’ argument. One could simply run through the argument in thought.” And
since Jackson hasn’t told us anything about how thoughts refer, Stich thinks the original
problem about the validity of the eliminativists’ argument still remains (p. 201).

Having worried about the difficulties involved in the concept of representation, we can
now move on to worry about a different concept, innateness. In her paper, Fiona Cowie
points out that the concept “innate” looks like the sort of concept that a philosopher like
Stich might at one time have wanted to throw out of cognitive science altogether. It
seems like a vague, contested concept, the many uses of which have little in common. So
Cowie, noting Stich’s skepticism about reference and representation, wonders why he
isn’t equally skeptical, or eliminativist, about innateness. Cowie makes a spirited defense
of the utility of messy concepts. Appealing to some fascinating episodes in the history of
science, Cowie argues that inquiry cannot proceed without some concepts. In the
absence of better, more precise concepts, we must struggle onwards, in a spirit of what
she calls vulgar pragmatism, with the ones we’ve got. We may know that our concepts
are imperfect, but the option of throwing everything out and starting again, in a kind of
philosophical Jacobinism, is not viable. Much twentieth-century philosophy of science
assumed that what we want is a much more precise language in which scientific claims
could be stated transparently. If Cowie is right – and Stich largely agrees with her – that
seems thoroughly wrongheaded, since some vagueness is both inescapable and perhaps
even desirable. Messy, vague concepts might promote fruitful inquiry, since concepts
defined too tightly, as Peirce suggested, would lock us into the current theories that
define the concepts, and make us unable to adapt our concepts to new evidence. Progress
would grind to a halt if our language was so precise that we had to rejig it every time we
found out something surprising.
Questions about the right way for inquiry to proceed raise issues in epistemology, where Stephen Stich has made important and provocative contributions. These can be divided into two coherent but clearly separable parts. The critical part is skeptical about whether the methods of analytic epistemology can deliver correct epistemological principles or theories. The positive part is a defense of pragmatism. The critical side of Stich’s epistemology saw its most sophisticated development in a paper he co-authored with Jonathan Weinberg and Shaun Nichols (WNS 2001). WNS dub the analytic philosopher’s method for testing and developing epistemological theories “Intuition Driven Romanticism.” It involves three features: (1) It takes epistemic intuitions (spontaneous judgments about the epistemic properties of cases) as inputs; (2) it produces epistemic claims or principles as output; and (3) the output is a function of the input – in the sense that significantly different inputs would yield significantly different outputs (432). WNS then proceed to conduct a number of empirical studies which suggest that there are significant and systematic differences in the epistemic concepts and judgments people in different cultures employ in evaluating cognition. In a typical study, WNS found that when given a Gettier-style example, a large majority of Western subjects gave the answer sanctioned by analytic epistemologists (“only believes”) but a majority of East Asians and a majority of subjects from India gave the opposite answer (“really knows”) (443). If these results are true, WNS conclude that the proponent of Intuition Driven Romanticism faces an unpalatable dilemma:

they must either argue that these intuitive differences . . . would not lead to different normative conclusions, or they must bite the bullet and argue that divergent normative claims are genuinely normative, and thus that the sorts of doxastic states that ought to be pursued by relatively rich and well-educated people are significantly different from the sorts of doxastic states that poor and less well educated folks should seek. (447–8)

Ernest Sosa calls into question the assumptions behind this criticism of analytic epistemology. According to Sosa, “the individual philosopher has intuitive access to data such as the Gettier examples, and can take these data into account in assessing criteria of rightness. Once he shares his thoughts with others, the philosopher may encounter apparent disagreement” (p. 105). Sosa offers three specific reasons for thinking that the disagreements discovered by WNS are merely apparent – they do not signal any deep disagreements about the nature and conditions of knowledge. First, subjects from different cultures might flesh out the thumbnail descriptions of (e.g., Gettier) cases differently. And so the different responses might be the result of different understandings of the case rather than different concepts of knowledge. Second, the disagreement might be an artifact of subjects not being given more choices. Rather than being offered the choices “Really Knows” or “Only Believes,” subjects might have been offered the choice of “Can’t Tell.” And third, the disagreement might be merely verbal – different subjects might simply mean different things by ‘knowledge’. But even if these considerations fully explain the cultural diversity, one might worry that Intuition Driven Romanticism cannot deliver a theory that carries out the Normative Project, which “attempts to establish norms to guide our epistemic efforts. Some of these norms may be explicitly
regulative, specifying which ways of going about the quest for knowledge should be
pursued and which should not” (WNS: 430). Sosa argues that theories of knowledge and
justification are not in the business of telling us what beliefs we ought to form or sustain.
Some beliefs might concern such insignificant matters that regardless of whether they
count as knowledge or are highly justified, we ought not believe them: “One should not
be forming any opinion, positive or negative, on that question. One has better things to
do with one’s time, even if we restrict ourselves to properly epistemic concerns” (p. 110).
And so there is no reason to postulate any deep cross-cultural epistemic disagreements.
Different cultures might evaluate beliefs according to different epistemic standards, and
this disagreement is no more problematic than one person rating a car with respect to
speed and another rating it with respect to gas mileage.

Stich agrees with Sosa that the empirical findings have not established cross-cultural
epistemic diversity “beyond a reasonable doubt” – although we might note that this is a
standard that is seldom met in philosophy. With respect to the three specific objections
to WNS’s empirical findings, Stich admits the first two (subjects understand the cases
differently, and there might be greater agreement with a “Can’t Tell” option) are genuine
possibilities which could be tested empirically. But with respect to the third possibility
(the disagreement may be merely verbal), Stich argues that we need a theory of content
that grounds the claim that people in different cultures might mean different things by
the term ‘knowledge’. To see this, suppose two people from different cultures fully
understand the non-epistemic facts concerning Smith’s belief that $p$ and still disagree
about whether that belief is an instance of knowledge. Some philosophers defend a
theory of content that implies that there can be no genuine disagreement here – there can
only be the application of different concepts (Jackson 1998; Pollock and Cruz 1999). But
Sosa agrees with Stich that it is possible that two people might have a genuine disagree-
ment about epistemic intuitions, e.g., a disagreement about the nature and conditions of
knowledge. Stich argues that the implications of such disagreements for philosophy are
radical. “For 2,500 years, philosophers have been relying on appeals to intuition. But the
plausibility of this entire tradition rests on an unsubstantiated, and until recently unacknow-
ledged, empirical hypothesis – the hypothesis that the philosophical intuitions of people
in different cultural groups do not disagree.” If this empirical hypothesis turns out to
be false, then a lot of philosophy “belongs in the rubbish bin” (p. 232). Finally, Stich
suggests that much of his debate with Sosa might rest on a fundamental disagreement
about what epistemology is supposed to do. Stich wants an epistemological theory that
has a robust prescriptive or recommendatory function – a theory that gives us guidance
about how to lead our cognitive lives. Sosa wants epistemological theories that tell us
about the nature and grounds of knowledge, justification, and other epistemic categories; he
doubts that these categories provide any useful guidance about how we ought (epistemically)
to lead our cognitive lives. So even if people have the same epistemic intuitions and even
if those intuitions give us direct access to the nature and conditions of knowledge, it
would still not be reasonable to appeal to those intuitions in constructing an epistemological
theory with real prescriptive bite. On that point, it would appear, Stich and Sosa agree.

Michael Bishop offers a generally sympathetic sketch of the development of
Stich’s epistemological thought. According to Bishop’s narrative, Stich’s epistemological
arguments drive us to the Quinean conclusion that genuinely normative epistemological theories spring not from a priori reflections on our intuitions but from the findings of empirical science. Bishop embraces this methodological result, but he balks at Stich’s pragmatism. On this view, “one system of cognitive mechanisms is preferable to another if, in using it, we are more likely to achieve those things that we intrinsically value” (1990: 24). Stich argues that pragmatism leads directly to relativism – the idea that the cognitive evaluation of a set of reasoning strategies will be highly sensitive to facts about the person (or group) using those strategies – for two reasons. First, there are likely to be significant differences in what different people intrinsically value, and so we should expect a pragmatic view to recommend quite different cognitive systems to different people (Stich 1990: 136). And second, a set of reasoning strategies that yields the best expected consequences in one environment might not do so in a different environment (Stich 1990: 136–40). The core of Bishop’s argument against Stich assumes two empirical hypotheses: (1) Our epistemic practices – the way we are naturally disposed to evaluate ways of forming beliefs – is implicitly reliabilist (truth-linked); (2) It would be very difficult, and perhaps practically impossible, for us to replace our reliabilist practices with purely pragmatic ones. These empirical suppositions would give us pragmatic reasons to adopt a reliabilist approach to cognitive evaluation. This is especially so if we want a theory that has a robust prescriptive function – that helps guide our cognitive endeavors.

In reply, Stich grants that if these hypotheses were true, they might give him pragmatic reasons to embrace reliabilism. But he notes that Bishop offers no empirical support for these hypotheses; they are merely speculations, and so “the verdict must be that Bishop has failed to make a convincing case for the pragmatic superiority of” reliabilism (p. 244). One of the more intriguing aspects of Stich’s reply is his discussion of the proper role of intuitions in epistemology. He distinguishes five different, though perhaps overlapping, epistemological projects where one might employ intuitions as evidence. Roughly, these projects are: (1) capturing people’s intuitions by developing a theory that entails them; (2) accounting for the tacit theory or competence underlying people’s intuitions; (3) analyzing people’s epistemic concepts; (4) characterizing the nature and conditions of knowledge, justification, etc.; (5) evaluating and prescribing ways of forming beliefs (pp. 237–8). Stich argues that intuitions are “an entirely appropriate source of evidence in the first three of these projects.” But he argues that the use of intuitions in the last two projects “is much more problematic” (p. 238).

The next set of essays in the volume turns from ontology and epistemology to the organization of the human mind. In his defense of eliminativism, Stich made an assumption that was standard in the early-to-mid 1980s: Folk psychology is a theory, known to all normal adults, that consists of nomological generalizations about mental states such as beliefs and desires. But philosophers and psychologists began to wonder whether this was the best way to understand the phenomenon of “mindreading” – the ability to attribute mental states to others, and explain and predict behavior on that basis. The debate looked clear cut: Mindreading was to be explained in terms of a theory of mind (“the theory-theory”) or as a form of simulation, in which the mindreader imaginatively puts herself in another person’s place so she can use her own cognitive and decision-making machinery to simulate that person’s mind. As the debate proceeded, things
changed. Many theorists arrived at various sorts of hybrid theory, which employed
the theory-theory to explain some aspects of mindreading and the simulation theory to
explain other aspects of mindreading. Alvin Goldman (2006) arrived at such a view, as
and Stich drew a further, more radical conclusion: there is no natural or theoretically
significant category of simulation. This is the thesis Alvin Goldman takes issue with in
his contribution. Goldman aims to show that mental simulation is “both a natural and
theoretically interesting category” that fits comfortably with widespread and progressive
forms of explanation in the cognitive sciences, particularly cognitive neuroscience. Citing
evidence from clinical, imaging, and single-cell recording studies, Goldman argues that
in typical cases, the same brain systems that are active when we are in a mental state or
perform an action are also active when we think about that state or action. These results
suggest that the category of simulation is unified at the neurological level, even if
simulation events have little in common phenomenologically or functionally. Goldman
argues that Nichols and Stich have failed to properly appreciate “the full power and
merit of the simulation theory” because they neglect cognitive neuroscience as an import-
ant source of insight in the investigation of mindreading.

Did *Mindreading* neglect relevant neuroscientific evidence? Stich follows Goldman
(2006: 113) in distinguishing between high-level mindreading (“the formation of beliefs
about propositional attitudes like beliefs, desires, intentions and decisions”) and
low-level mindreading (“the formation of beliefs about sensations, like feelings of pain,
and about emotions, like fear, disgust and anger”). *Mindreading* was explicitly about the
former. Since the neuroscientific evidence Goldman cites that was available prior to 2003
(when *Mindreading* appeared) is directly relevant only to low-level mindreading, Stich
argues that he and Nichols did not neglect any relevant neuroscientific evidence. As for
the substantive issue – does simulation identify a unified, natural category? – Goldman
proposes that mental simulation is a mental process that is executed with the (often
unconscious) aim of duplicating or matching another mental process in some respect
(2006: 38). The problem with this account, as Stich sees it, is that it too fails to identify
a natural category. Consider two cases of mindreading that engage the same cognitive
processes and that involve ascribing the same propositional attitudes – the only differ-
ence being that the attributions are in one case correct and in the other case incorrect.
Stich argues that on Goldman’s view, the former case counts as simulation but not the
latter. And that’s because the former case resembles the target process whereas the latter
does not. Stich admits that Goldman might argue that the latter case is an episode of
simulation because the aim of the executed process was to duplicate the target process.
But then Goldman needs to patch his view with a general account of the unconscious
aims and purposes that drive simulation and some reason to think that these unconscious
aims are typically present, and Stich does not expect such a patch to be forthcoming.

Mindreading and nativism, the issues addressed by Cowie and Goldman, come together
in Kim Sterelny’s contribution. In *Mindreading*, Nichols and Stich argue that our
mindreading abilities are subserved by an array of systems that employ both theory and
simulation; these include perceptual, reasoning, and planning systems that have evolved
for other purposes and been co-opted in the service of attributing propositional attitudes.
Nichols and Stich studiously avoided taking a position on whether any part of these mindreading cognitive systems is innate. But Sterelny argues that their hybrid model significantly weakens the case for nativism.

The two forms of nativism Sterelny considers are both supported by variants of the same argument. Mindreading, like language, develops in nearly all humans (a) early, (b) independent of most forms of learning deficit, and (c) regardless of cultural variation and informational input from the environment. Therefore, some significant aspect of the cognitive machinery responsible for mindreading cannot be learned, and so must be innate. Concept nativism holds that intentional concepts, such as belief and desire, are innate. This is because intentional concepts refer to unobservable, inner states of a subject. And since our mindreading abilities are early and robust, these concepts must not be learned but innate. Sterelny argues that the hybrid model coheres naturally with a non-nativism about intentional concepts. On the hybrid model, mindreading is not explained entirely in terms of theoretical knowledge – simulation-type processes do some of the mindreading work. And so our standard cognitive endowment might not come with built-in intentional concepts. These might be the result of learning. The second form of nativism Sterelny considers is informational nativism, which holds that the information that we need for mindreading is innate. On the hybrid model, the modules on which early mindreading draws are shallow perceptual systems without a rich database. Sterelny argues that as our mindreading abilities develop and become more sophisticated, the information recruited for mindreading is provided by one’s cultural “scaffolding” and especially one’s linguistic environment.

With respect to concept nativism, Stich grants that Sterelny’s account of how we acquire intentional concepts might be true. But much is still obscure: What are concepts? What is the mechanism Sterelny thinks is responsible for producing our intentional concepts? And how are we to distinguish which concept acquisition mechanisms are nativist (and don’t involve learning) and which are empiricist (and do involve learning)? Too much is still in darkness for Stich to renounce his studied neutrality. As for informational nativism, Stich agrees that Sterelny is right: After about the age of three, familiar processes of learning are all we need to explain the increased sophistication of our mindreading abilities. But Stich is skeptical that anyone – or at least anyone “worth refuting” – holds the informational nativist view that Sterelny attacks.

The papers in this volume cover much of Stich’s work over the last 25 years, and with Jesse Prinz’s contribution we arrive at one of Stich’s most recent concerns, moral psychology. Prinz criticizes a number of nativist views about moral psychology, including a view about the nature of norms defended by Sripada and Stich (2006). A norm, for Sripada and Stich, “is a rule or principle that specifies actions which are required, permissible or forbidden independently of any legal or social institution” (281). There are norms in all cultures, and although there are certain classes of norms that are found in all cultures (most societies have norms against killing and incest, and norms that foster sharing and cooperation), the specific content of norms exhibit great variability in different groups (282–4). What’s more, norms often permit specific kinds of exceptions (284). Sripada and Stich propose an account of the psychological mechanisms subserving norms. They posit “two closely linked innate mechanisms, one responsible for norm
acquisition, the other for norm implementation” (289). Prinz argues that morality is an “accident” – a by-product of the existence of cognitive systems that have evolved for purposes other than norm acquisition. While there are many routes to the acquisition of norms, specifically moral norms are just a species of more general norms enforced by emotions of disapproval. Prinz grants that the universal tendency for humans to develop the emotion-backed norms he identifies with moral norms is the result of our biological endowment. However, he does not count that as a species of moral nativism. That’s because no part of the biological endowment that is responsible for the acquisition of our moral norms is specifically adapted for acquiring moral norms.

Stich’s response does not address the nativism issue in any detail. Rather, he focuses on a difficulty in Prinz’s paper that he thinks is endemic to a lot of recent empirically informed work in moral psychology. Stich wants to know: What is an empirically informed theory about the psychological underpinnings of morality trying to do? In particular, what makes a theory of norms (or rules or judgments) a theory of moral norms (or moral rules or moral judgments)? One possibility is that which norms count as moral norms are determined by our concepts or intuitions about moral norms. Another possibility assumes that moral norms constitute a natural kind and the theory is aiming to provide an account of that natural kind. The problem, according to Stich, is that there is no guarantee that either project can succeed. What’s more, Stich thinks that Prinz’s theory fails on both of these projects – and he suggests that Prinz’s theory is not alone in this regard. With respect to the first project, there are norms that count as moral norms according to Prinz’s theory (i.e., violations of norms of about eating, ritual behavior, and the disposal of corpses can trigger strong emotions of disapproval) which are intuitively not moral norms. And with respect to the second project, Prinz’s theory does not seem to capture a natural kind. Prinz takes moral norms to be “accidents” whose causes and content can vary dramatically across cultures, and “[a]ccidents don’t make good natural kinds” (p. 225). Stich is calling on those who are committed to putting some real psychology into moral psychology to reflect on what exactly they want their theory to do. For the record, Stich thinks that the view of norms he and Sripada defend escapes this worry. That’s because they argue that norms – but not moral norms – are a natural kind of social science.

References


1

Is There a Role for Representational Content in Scientific Psychology?

FRANCES EGAN

Steve Stich used to be an eliminativist. As far as I can tell, he renounced eliminativism about the time that he moved from the west to the east pole.1 Stich was right to reject eliminativism, though I am not convinced that he rejected it for the right reasons. Stich 1983 contains a comprehensive attack on representational content, a central feature of both folk psychology and the Representational Theory of Mind, the leading philosophical construal of scientific psychology. Stich’s current position on the role of content in psychological explanation is not entirely clear. One of my aims in this chapter is simply to invite Stich to clarify his views on representational content; the question that forms the title of this chapter is therefore addressed directly to Stich. I begin by sketching his original anti-content argument. I then trace some later developments in his thinking about content. I argue that content does play an important role in scientific psychology, for precisely the reasons that Stich identified in his original argument against content. I conclude with some general remarks on eliminativism.

1

Stich characterizes four distinct claims that eliminativists are inclined to make:

1 ‘Belief’, ‘desire’, and other familiar intentional-state expressions are among the theoretical terms of a commonsense theory of the mind. This theory is often called folk psychology.
2 Folk psychology is a seriously mistaken theory. Many of the claims it makes about the states and processes that give rise to behavior, and many of the presuppositions of these claims, are false.
3 A mature science that explains how the mind/brain works and how it produces the behavior we observe will not refer to the commonsense intentional states and processes invoked by folk psychology. Beliefs, desires, and the rest will not be part of the ontology of a mature scientific psychology.
4 The intentional states of commonsense psychology do not exist. (Stich and Ravenscroft 1994: 116)
Claim (1) is simply a presupposition of eliminativism. Realists about beliefs and desires typically endorse it too. We will call claim (3) weak eliminativism (following Stich 1992: 245), and a commitment to either (2) or (4) strong eliminativism. (2) and (4) are, of course, distinct claims, but we will note their differences only where relevant. Stich 1983 endorsed all three eliminativist claims. He also assumed there that (3) and (4) follow from (2).

Stich’s 1983 argument for eliminativism was two-pronged. In the latter part of the book, he argued that folk psychology makes substantial assumptions about the cognitive architecture underlying our cognitive capacities and behavioral repertoire. He claimed, for example, that folk psychology is committed to the idea that a single memory system suberves verbal and non-verbal behavior (p. 231), and that belief organization and storage is modular. There is good reason, the argument continues, to suspect that these assumptions are false. I argue in Egan 1995a that folk psychology involves no substantive commitments about architecture or cognitive processing, and so the falsity of these architectural assumptions would not threaten it. I shall not defend this claim here, though the issue will come up again in the last section.

My concern here is with the other prong of Stich’s 1983 argument for eliminativism, that is, his case against representational content. First, some stage-setting. Folk psychological predictions and explanations of behavior appeal to content-specific beliefs and desires. For example, it is my belief that there is beer in the refrigerator together with the content-appropriate desire (to drink a beer, or perhaps just to drink something cold), that explains my going into the kitchen and getting a beer. Appealing to my belief that there is beer at the local bar or my desire to win the lottery fails to provide an explanation of my beer-fetching behavior. Moreover, this behavior is rational just to the extent that it is caused by content-appropriate beliefs and desires.

The case against content in Stich 1983 is a tour de force. Stich argues persuasively that content ascriptions are both vague and context-sensitive. For any given predicate of the form 'believes that p' there will be many contexts where there is simply no saying whether it applies or not, and hence it will often be unclear whether a generalization that invokes such a predicate applies to a given subject. Appeals to content are also observer-relative. As Stich puts it, “To believe that p is to be in a belief state similar to the one underlying our own sincere assertion of ‘p.’” (136). Moreover, appeals to content often presuppose both ideological similarity and reference similarity. Two beliefs are ideologically similar if and only if they are embedded in similar doxastic networks. Suppose, for example, that two subjects both say “Senator Smith is a liberal.” Whether or not we would be inclined to attribute the same belief – that is, a belief with the same content – to the two subjects depends on whether their other beliefs involving the concept liberal are similar. Two beliefs are reference similar if and only if the terms subjects use to express the beliefs have the same referent. Subjects using the vocable ‘granite’ to refer to different substances would express different beliefs when they say “granite counter-tops are durable.”

Let us call these various properties of content – vagueness, context-sensitivity, observer-relativity, the tendency to presuppose ideological and reference similarity – the R (for ‘relativity’) properties. The problem, Stich argues, is that a taxonomy that has the R properties will impose a more fine-grained individuative scheme than is appropriate for use in scientific psychology. In particular, a taxonomy that has the R properties will not respect the Autonomy Principle, which holds that “any state or property properly invoked in a
psychological explanation should supervene on the current, internal, physical state of the organism. Stich argues (1983: 167–9) that systematic explanations of an organism’s behavior of the sort that psychologists seek to provide – as opposed to those sought by, say, social historians or biographers – will apply equally to an organism’s physical duplicate. Such explanations should invoke only narrow states and properties shared by all duplicates; in particular, they should invoke only narrow causal role. But states individuated in part by their content, as beliefs and desires are, build in various features of the subject’s historical, environmental, and social context, features that are irrelevant for scientific psychology.

Stich offers what he calls the ‘replacement argument’ in support of the autonomy principle (1983: 165). Since physical duplicates would behave identically in all contexts, and since psychology is the science that aspires to explain behavior, any states or properties not shared by duplicates must be irrelevant to psychology. Of course, there are some behavioral descriptions that do not apply equally to duplicates. I can sell you my car; a physical duplicate of me cannot. Stich proposes that psychology restrict itself to ‘autonomous behavior descriptions’, that is, descriptions that satisfy the following condition: if they apply to an organism in a given setting then they would apply to any replica of the organism in that setting. Scientific psychology should aim to purge itself of language that builds in historical or contextual features. It should purge itself, therefore, of representational content. So, to summarize: one important strand in Stich’s 1983 case for eliminativism, which we will call the 1983 anti-content argument, is the following:

1. Content has the \( R \) properties. (It is vague, context-sensitive, presupposes various dimensions of similarity, etc.)
2. States and properties invoked in a scientific psychology should supervene on the current physical state of the organism (the Autonomy Principle).
3. Properties with the \( R \) properties do not supervene on the current physical state of the organism, and so a taxonomy based on such properties will violate the Autonomy Principle.
4. Therefore, content should not be invoked in a scientific psychology.

II

Stich’s original argument concerned ordinary content ascribed in folk psychological predictions and explanations of behavior. In Stich 1991, he argued that narrow content – content that prescinds from the subject’s historical, environmental, and social context, and hence is shared by physical duplicates – is nonetheless still too vague and context-sensitive (in other words, it still has many of the \( R \) properties) to serve in a scientific psychology. In particular, he argued, narrow content is ill-suited to play a role in computational models of mind. Such models individuate mental states in terms of their narrow causal role but, as numerous examples show, narrow content does not track narrow causal role. Summing up, Stich says “the categories of a narrow content taxonomy are simply the categories of a broad content taxonomy extended to meet the demands of the principle of autonomy. But the broad content taxonomy of commonsense psychology is too vague, too context-sensitive and too unstable to use in a serious scientific theory. Narrow content
inherits all of these deficits” (1991: 250). Clearly, at this point, Stich still believed that content, broad or narrow, is not suitable for science. And he still endorsed weak eliminativism – scientific psychology will not invoke commonsense intentional states because these states are individuated by content. The argument, which does not rely on the Autonomy Principle, looks something like this:

1. Content has the \( R^* \) properties. (Where \( R^* \) properties are the proper subset of the \( R \) properties that meet the demands of the Autonomy Principle, including, vagueness, context-sensitivity, and instability.)
2. Properties with the \( R^* \) properties (vagueness, context-sensitivity, etc.) are not suitable for science.
3. Therefore, content should not be invoked in a scientific psychology.

We will call this the 1991 anti-content argument.

By 1996 Stich’s view has changed considerably. In 1996c, he says “being invoked in a successful science is all that it takes to render a property scientifically legitimate. On my view, the jury is still out on the question of whether successful science can be constructed using intentional categories” (1996c: 199). But one would have thought that the jury had already handed down its verdict. Has content somehow been rehabilitated? Let’s consider some recent developments in Stich’s thinking that might account for his change of mind:

(1) The so-called “naturalization project” is the attempt to specify, in a non-intentional and non-semantic vocabulary, sufficient conditions for a mental state’s meaning what it does. Most, if not all, attempts at naturalization have failed to meet these stringent requirements while characterizing something that looks sufficiently like representational content. Stich argues persuasively (in Stich 1992 and Stich and Lawrence 1994) that the failure of the naturalization project would not impugn intentional content. Content, whatever its other failings, does not need to be naturalized. But this conclusion should give Stich no reason to reconsider whether content is fit for use in scientific psychology. His case against content – in particular, the 1983 and 1991 anti-content arguments – does not depend upon metaphysical or philosophical considerations of the sort that vex those engaged in the naturalization project.

(2) Stich has become convinced that even if folk psychology is a seriously mistaken theory (claim 2 on p. 14), it does not follow either that scientific psychology will find no use for intentional categories (claim 3), or that beliefs and desires do not exist (claim 4). To think otherwise, as Stich once did, he argues in 1996b, is to make some dubious assumptions about the reference of theoretical terms. But the failure of the inference from claim (1) to (3) and (4) does not undermine the 1983 or the 1991 anti-content arguments, neither of which depends upon any assumptions about reference. The failure of the inference provides no reason to doubt that content has the \( R \)-properties, and so no reason to be optimistic that successful psychology will invoke intentional states.

(3) The 1991 anti-content argument does not depend upon the Autonomy Principle. Still, it is worth noting that Stich is no longer willing to endorse the principle. In both 1991 and 1996b he holds the Autonomy Principle at arm’s length, claiming:
There is, to put it mildly, considerable controversy surrounding this thesis. Some writers, myself included (but I was younger and much more naive at the time), have claimed that it is intuitively obvious . . . Others have tried to defend the thesis by deducing it from other, perhaps less controversial, metaphysical doctrines; still others have claimed that it is simply false. (1996b: 23)

He goes on to say: “But if it is not clear whether the thesis is defensible, it is clear that if the thesis is accepted then . . . folk psychology is in trouble” (1996b: 23). The Autonomy Principle requires that psychological properties supervene on the current physical state of the subject, and so it clearly prohibits individuating psychological states in terms of their broad content, as folk psychology appears to do. Rejecting the Autonomy Principle would let broad content back into the picture, if there were not independent reason to think that content – broad or narrow – is also vague, context-sensitive, and unstable. There is no obvious reason for Stich to repudiate his 1991 anti-content argument, which relies on the claim that content has the \( R^* \) properties and hence is not suitable for use in a scientific psychology.⁸

In short, recent developments in Stich’s thinking do not account for his apparent change of mind about content. I will wrap up this section with a question for Stich: are there (still) principled reasons why scientific psychology should avoid a commitment to representational content?

III

Here is what I will argue:

1. Content has the \( R \) properties.
2. A qualified version of the Autonomy Principle is true. In particular, a theory concerned primarily with characterizing the mechanisms underlying our cognitive capacities should employ taxonomic principles that prescind from the mechanism’s normal context.
3. Therefore, a theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should not individuate by properties with the \( R \) properties (e.g., content) – it will violate the Qualified Autonomy Principle.
4. The theories that are in the business of characterizing the mechanisms underlying our cognitive capacities – computational cognitive theories – in fact, do not individuate by content.
5. Content nonetheless plays an important explanatory role in such theories, and it is able to play this role in part because it has the \( R \) properties.

I will not defend (1); I take it to have been established by Stich 1983. Content is vague, observer-relative, context-sensitive; it presupposes reference and ideological similarity. My argument for claim (4) will be quite brief, as I have defended it at length elsewhere. Turning, then, to the argument for (2). Recall that the Autonomy Principle holds that the only properties to be invoked in a scientific psychology are \( narrow \) properties that supervene on the current physical state of the organism. When the principle is appropri-
ately qualified it has a compelling rationale. If a theory is concerned primarily to characterize the mechanisms and processes underlying the behavior and capacities of a complex system, then relatively narrow taxonomies are better for the following reason: the narrower the individuative scheme—that is, the greater the range of contextual properties it prescinds from—the greater the scope of the theory’s generalizations. **Narrow taxonomies maximize generality.** Hence, the **Qualified Autonomy Principle** (henceforth, QAP) holds that a psychological theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should employ taxonomic principles that prescind from features of the mechanism’s normal context.

Consider the inhabitants of Twin Earth. According to Putnam’s (1975) myth, their cognitive capacities and dispositions to behavior are the same as ours. Characterizing the underlying commonalities between ourselves and Twin Earthlings, rather than obscuring such commonalities by building contextual features into our taxonomies, provides a basis for explaining and predicting our behavior, and theirs, in a wide range of counterfactual circumstances. This strategy does not ignore or downplay the subject’s environment as a determinant of her behavior. It simply requires that environmental and other contextual determinants of behavior be specified as independent variables.

QAP is in one sense stronger than Stich’s original principle. The states and properties invoked by theories that purport to characterize the mechanisms responsible for cognitive capacities will typically supervene not just on the current physical state of the whole subject, but more narrowly on the current physical state of the mechanism itself. The theorist is directed to characterize the relevant mechanisms and processes independently of the larger systems in which they are embedded, prescinding from details of both external and **internal** (i.e., intra-organism) environment.9 The resulting theories will be **radically internalist**; the boundary of the subject’s skin will have no particular **individuative** significance.10

QAP is a special case of a more general principle about how to understand and explain complex systems. As noted above, taxonomies that prescind from contextual features, requiring that they be specified as independent parameters, provide a basis for predicting and explaining the system’s behavior in a wide range of counterfactual circumstances. The presumption in favor of narrow taxonomies applies only to theories in the business of explaining how complex systems work. It does not apply to “historical” theories, such as evolutionary biology and geology, whose explanatory goals are somewhat different.11 Thus, a theory that aimed to explain how circulation is possible might subsume human and Martian hearts under the same kind if they worked the same way, whereas evolutionary biology, with an interest in the specific origins of biological mechanisms on earth, type-identifies only homologous organs.

In another sense, then, QAP is weaker than Stich’s Autonomy Principle. It does not require scientific psychology to restrict itself to narrow taxonomies. Not all of psychology is concerned to characterize the mechanisms underlying cognitive capacities and behavior. Computational psychology clearly is, but other branches of scientific psychology have other goals. Developmental psychology, for example, is concerned to characterize the particular stages in a human child’s cognitive development, and so we would not expect it to respect QAP.

It follows from (1) and (2) that a theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should not individuate psychological