Semantic Externalism and Knowing Our Own Minds: Ignoring Twin-Earth and Doing Naturalistic Philosophy

by

RICHARD BOYD

Cornell University

Abstract: In this article I offer a naturalistic defence of semantic externalism. I argue against the following: (1) arguments for externalism rest mainly on conceptual analysis; (2) the community conceptual norms relevant to individuation of propositional attitudes are quasi-analytic; (3) externalism raises serious questions about knowledge of propositional attitudes; and (4) externalism might be OK for "folk psychology" but not for cognitive science. The naturalist alternatives are as follows. (1) Community norms are not anything like *a priori*; sometimes they are incoherent. (2) Often propositional attitudes lack determinate content: we do not know the content of thoughts or sentences because there is no fully definite content to be known. (3) Often achieving determinate content is a major socially mediated cognitive achievement that depends on just the factors of social and environmental embedding posited as individuative by externalists, so (4) externalism explains how people can, sometimes, come to have, and to know, determinate attitude contents. (5) Reference and content, for both thought and language, are determined by complex and messy dialectical relations involving many such environmental and social factors; consequently, determinate reference, truth-conditions, etc., are somewhat uncommon outcomes. (6) The basic semantic relation is (typically imperfect) socially mediated accommodation between perceptual, cognitive, linguistic, classificatory and inferential dispositions and relevant causal structures in the environment. (7) This accommodation explains how concepts, language, taxonomies, etc., contribute to individuals' rational inductive, explanatory and practical achievements. (8) So externally individuated propositional attitudes are required for cognitive science explanations of individual human rationality and its inductive and explanatory achievements. "Individual rationality ain't (entirely) in the individual head."

Keywords: semantic externalism, propositional attitudes, reference, accommodation, natural kinds, projectibility, norms, malignant meanings, naturalism

1. Introduction

1.1 Semantic externalism and knowledge

WHEN WE THINK AND TALK we do so in contexts determined both by referentially relevant features of our environment (the stuff we think or talk about) and by (another special case of our environment) the linguistic, conceptual and inferential norms of the communities (NB: plural) to which we belong. Ever since the initial work of Kripke (1980) and Putnam (1975) on "causal" theories of reference many philosophers have insisted that the propositional contents (meanings) of sentences are determined by the first of these external factors as well as by the second. The pioneering work of Putnam (1975) and of Burge (1979, 1986) has lent credibility

to the claim that the contents of our individual propositional attitudes should also be individuated by external factors of one or both sorts.

Unsurprisingly, semantic externalist conceptions of content raise interesting epistemological questions, both about the epistemic status of semantic externalism and about its epistemic implications. Extraordinarily sophisticated philosophical arguments have been directed toward addressing these questions. Three features of the literature interest me here. First, very often philosophers' positions are under-written by very nearly *a priori* arguments about exceptionally cleverly contrived logically possible examples.

Second, there is an assumption about the relationship between the adequacy of a person's linguistic or conceptual understanding and the extent to which her think-ing conforms to her community's conceptual or linguistic norms. Roughly, com-munity norms are treated as though they are something like "analytic at a time" (*sensu* Putnam, 1962) or otherwise privileged as standards of cognitive or linguistic competence, so that to attribute to someone a belief that flies in the face of such norms is, *prima facie*, to attribute to her some inadequacy or incompleteness of conceptual understanding (and, perhaps, to raise the question of whether the attribution meets an appropriate standard of "charity" in such matters).

Finally, a central issue in the literature has been the extent to which adopting (one or another version of) externalism about the content of propositional attitudes might commit one to implausible conclusions about the extent to which people can fail to know the contents of their own beliefs and other propositional attitudes (see, e.g., Brown, 2004; Burge, 1988, 1996; Gibbons, 2001; McLaughlin and Tye, 1998; Nuccetelli, 2003; Sawyer, 2002; Segal, 2003; Wikforss, 2001, 2004; Wright, Smith and Macdonald, 1998).

I have worries about each of these features of the literature. I think that the best case(s) for externalism (and probably the best cases against) rest on consideration of more complicated real-life cases of propositional attitude attributions and of their explanatory functions. I think that pretty often the prevailing conceptual and linguistic norms in a community (I will focus here mainly on scientific communi-ties) are anything but analytic at a time. Pretty often they are a mess and sometimes they are deeply incoherent. Not uncommonly adequacy of conceptual understand-ing can only be achieved at the expense of community norms. Finally, I think that most of the time people really do not know the propositional content of their propositional attitudes, in part because much of the time their propositional atti-tudes do not have determinate content for them to know. So, in so far as some version of externalism about the content of propositional attitudes entails a very weak conception of first person authority, that would by my lights *prima facie* count in its favour.

My aim here is to offer a distinctly naturalistic alternative to understanding and defending externalism and to explore the question of whether or not the

appropriateness of individuating propositional attitudes widely extends beyond "folk psychology" to "cognitive science." So I will challenge four fairly widespread and plausible views about externalism, to wit:

1)   Arguments for (or against?) externalist principles re individuation of propositional attitudes properly rest (mainly) on (all but) *a priori* conceptual considerations about appropriate attitude ascriptions with respect to philosophically illuminating imaginary cases.
2)   The community semantic/inferential norms which might figure in the individuation of attitudes are (something like) quasi-analytic.
3)   Semantic externalism re propositional attitudes raises serious problems, perhaps non-fatal, about subjects' knowledge, understanding of own propositional attitudes and of their truth conditions.
4)   Semantic externalism might be right for folk psychology but it is incompatible with the individualistic individuation of propositional attitudes required for scientific psychology/"cognitive science."

1.2 Alternative conception, summary: semantics and content
Here are the key ideas I will develop about the semantics and epistemology of content attributions.

1)   Community conceptual norms are not anything like *a priori* or *analytic at a time*; sometimes they are incoherent.
2)   Very often our propositional attitudes (and correlated sentences) lack determinate propositional content. I do not just mean that some chemist's belief might be equally well described as (on the one hand) the belief that a Bronsted acid is a proton donor and as (on the other) the belief that a Bronsted acid is a "proton-kindred proton-stage" donor. I mean that often we do not know what we are believing/talking about because there is no definite thing (even modulo philosophers' contrivances) that we are having beliefs/conversations about *even though our beliefs and conversations have real, causal, and explanatory correspondence referential relations to some subject matters*.[1]
3)   In such cases eventually arriving at a determinate content is a major socially mediated cognitive achievement which depends on a subject's being embedded in just the sorts of social and environmental factors cited in externalist conceptions of content, so

_____

1   Thanks to Rebecca Copenhaver and Eddie Cushman for helping me see that I should make it clear early on that I am not here trying to make a Quinean or Davidsonian point or raising issues about ontological relativity.

4)   externalism explains how in such cases people can, sometimes, come to have, and to know, determinate propositional attitude contents.

5)   Big semantic point: Reference and other key semantic relations are process-like dialectical relations such that determinate reference, truth-conditions, etc., are somewhat uncommon outcomes. Instead,

6)   the basic semantic relation is (typically imperfect) *accommodation* between perceptual, cognitive, linguistic, classificatory and inferential dispositions and relevant causal structures in the subject's environment which helps to explain how language, taxonomies and discourse can contribute to rational inductive, explanatory and practical achievements.

7)   In consequence externally individuated propositional attitudes are required for scientific psychological explanations of human rationality and human inductive and explanatory achievements.

1.3 Alternative conception, summary: content and psychological explanations
One could (and many people do) think that in everyday life, and perhaps in "folk psychology," we harmlessly individuate propositional attitudes widely (perhaps with respect to both environmental factors and community norms), while still thinking that such attitudes must be individuated narrowly for the purposes of scientific psychology or "cognitive science" (see Egan, 1995, for an especially clear articulation of this position). I think that this is mistaken, but my response is more speculative than that sketched in section 0.1. What I will outline here is what I believe to be the best *a posteriori* case for wide individuation in scientific psychology.

1)   Propositional attitudes attributions (help to) explain not just behaviours but also deliberative rational agency *and its successes and failures*.

2)   The relevant sort of explanation is (something like) rationalizing explanation, but rationalizing explanations are causal explanations positing the operation of people's rational capacities as causal factors.

3)   The relevant sort of rationality is not ideal rationality but actual deliberative rationality as exhibited by individual agents participating in norm-governed communities (NB: plural).

4)   How should we individuate attitudes? By causal powers/dispositions that explain their roles in individual deliberative rationality.

5)   What roles are the relevant ones? Here is the *rational discourse modelling thesis*: Those roles are approximately the same as the causal roles played in deliberation by associated *sentences* in relevant communities' (NB: plural) discourse(s).

6)   Why does this work?

Weak hypothesis: People internalize, as components of their internal inferential architectures, the inferential and semantic norms of the relevant communities and approximately implement them with respect to their own propositional attitudes, perhaps sometimes deploying linguistic resources in private (*internalized*) thinking. Think: G.H. Mead.

Externalist hypothesis: An individual's deliberative rationality ordinarily depends on continued embedding in relevant deliberative communities and the environment. Propositional attitude explanations presuppose (what are in fact) ongoing norm-implementing effects of social and environmental embedding.

While not denying the role of the internal factors posited in the weak hypothesis, I will defend the externalist hypothesis and thus the role of externalist factors in scientific as well as in everyday psychological explanations.

1.4 Cognitive science paraphrase

Philosophers sometimes seem so worried about being scientifically up-to-date that they risk falling into reductionist scientism. Nevertheless it is worth checking to see whether the case for *scientific* wide individuation can be formulated in fashionable cognitive science terms. OK, let us then be *oh so scientific*. Here is an outline of the cognitive science case.

1) Propositional attitudes are mental states.
2) Mental states are individuated by their role as computational states – as information processing states.
3) In the case of non-living computers the computational role of a state in one (physically bounded) computer can be partly realized in (and partly individuated by) information processing that partly takes place in some other distant (physically bounded) computer. (Think: networked computation; ATMs.)
4) Externalist hypothesis: Human deliberative rationality is social-and-environmental-network realized. We are networked computers. Roughly, individual rationality = participation in discourse communities + environment, so that
5) The laws, reliable generalizations, inference strategies, etc., *regarding individual rational behaviours, choices, deliberations, etc.*, upon which the reliability of belief-desire-rationality involving cognitive psychological explanations depend are true (approximately true, reliable, . . .) only because the *individuals* in question function in normatively regulated, environmentally embedded discourse *communities*. Some of each individual's individual rationality is realized in (and thus properly individuated in terms of) rational inferential processes in other people's heads and in communities' rational discourses where these are to be understood as including epistemically relevant causal interactions with the relevant subject matter(s).

6)   Old slogan: "Meaning just ain't in the head." New: "Individual rationality isn't (just) in the (individual's) head." Alternative formulation:

7)   "From each according to her (cognitive) abilities, to each according to her (cognitive) needs." (Actually, I will accept the unmodified slogan too, but that is beyond the scope of this article.)

2. On to the Arguments: Problems Regarding Understanding Our Own Sentences, Theories, Propositional Attitudes

2.1 Incomplete understanding, partial denotation, inadequate – even pathological – conceptual resources

Sentences first. It is utterly routine in scientific investigations and other sorts of serious inquiry that there are terms (and consequently sentences, doctrines, etc.) such that they lack (or at one time lacked) clear or determinate reference or meaning or truth conditions and regarding whose appropriate use (if any) we have (or had) profoundly inadequate conceptual understanding. Consider, for example(s):

> "species," "homology," "gene"
> "atom," "element," "valence"
> "knowledge," "truth," "agency"
> "gene"
> "instinct"
> "race"
> "sexual deviance"

Propositional attitudes next. Whatever non-conceptual representations there may be, you cannot have beliefs, or conjectures or questions about species, homology, genes, atoms, elements, valence, knowledge, truth, agency, or other esoteric subjects without deploying conceptual resources linked to such relevant linguistic items, so there is (or was) a corresponding lack of determinate meanings and truth condition and a profoundly inadequate conceptual understanding in the case of those propositional attitudes.

2.2 Malignant inferential norms (preliminaries)

In addition to the sorts of referential and conceptual deficiencies just mentioned there are closely related cases in which the *meaning-constitutive (!)* inferential norms within a scientific community fail quite pathologically to be truth-conducive. I will give examples below, but here are some preliminary definitions.

1)   By the *conceptual meaning* of term q/sentence s in a research community, C at a time, t, let us understand the cognitive/inferential commitment with

which one must *engage* in order to understand the use of q/s in literature and practice of C, at t. Let us recognize two different sorts of engagement as follows.

2)   By *uncritical engagement* let us understand the sort of *acceptance* of those cognitive/inferential commitments which characterizes the approach of an ordinary well-trained sophisticated practitioner in C at t. Uncritical engagement (not under that description of course) is what you go to graduate school in C's discipline to learn.

3)   By *critical engagement* let us understand the special sort of appreciation of the conceptual meanings in question which a philosopher, or historian or sociologist studying the practices of C at t might achieve *without necessarily accepting the relevant cognitive/inferential commitments.* Critical engagement (perhaps under something like that description) is what you go to graduate school in comparative religion to acquire with respect to more than one religious community (often) at more than one time.

With these definitions in hand, I can now state my central claim about conceptual meanings: There are serious scientific disciplines such that the conceptual meanings of their terms and key sentence types are such that uncritical engagement with them – the sort of engagement that goes with accepting and participating in the communities' epistemic norms – are profoundly flawed. An example follows.

2.3 Malignant conceptual meanings: human sociobiology/ *"evolutionary psychology"*

Recent work in "evolutionary psychology" illustrates the role of malignant meanings in contemporary science. (I will summarize the case briefly. For more details, see Boyd, 2001a, 2010; for other critiques see, e.g., Kitcher, 1985; Buller, 2005; Richardson, 2007; for a spirited defence, see Alcock, 2001.) To a good approximation the central inferential patterns in evolutionary psychology involve (1) advocating an evolutionary scenario, S, regarding selection for a behaviour, B, in the environment of evolutionary adaptation and (2) taking that scenario to "predict" that humans have an innate and relatively non-malleable unconscious motive with the same propositional content as the evolutionary function which S assigns to B. (For a better approximation add some inference patterns which trade on conflating the psychological use of "altruism" and "altruistic" with technical metaphorical uses of those terms in evolutionary theory. For an almost perfect approximation, add inferences from premises of the form "B has a biological/ genetic basis" to "B is innate and relatively non-malleable.")

These inference patterns reflect deep confusions about the evolution of behavioural repertoires and about the relationship between evolved behaviours and learning. No evolutionary biologist would admit to accepting them if they were made

explicit (see Alcock, 2001, for repeated denials that contemporary human socio-biology has methodological commitments like these). Nevertheless, one cannot understand the literatures – one cannot see what inferential connections are being taken for granted – unless one engages with these pathologically defective inference patterns. They are malignant. Human evolutionary psychology studies, among other things, human mate choice, child rearing, child abuse, altruism, cooperation and competition. All of the terms "human mate choice", "child rearing", "child abuse", "altruism", "cooperation" and "competition", as they are used in the socio-biological literature, refer to real aspects of human psychology or behaviour, but they do so despite, rather than because of, the inferential strategies associated with their conceptual meanings in sociobiology.

2.4 General lesson: knowledge, understanding of sentences, theories, propositional attitudes, their truth conditions

The case of malignant meanings, together with the examples cited in section 1.0, illustrate the point that the following are utterly routine in the sciences (and, of course, in other disciplines like philosophy as well).

1) Sentence types and corresponding propositional attitudes for which there are no determinate subjects, contents or truth conditions, and
2) seriously incomplete, or even profoundly defective, understanding of the truth conditions (for sentences and corresponding propositional attitudes) and of their appropriate inferential roles, where
3) these conditions can persist for decades.

Moreover, it is not uncommon (as the examples from evolutionary psychology illustrate) for one sort of scientific competence – the ability to read and understand the literature in a discipline – to be achieved at the expense of an equally important sort of competence – having a genuinely coherent conception of the relevant subject matter. This sort of situation, too, can last for decades. Indeed, meaning-constitutive confusions about "nature and nurture," of which the malignant meanings in evolutionary psychology are instances, seem extremely persistent. In order to assess the relevance of all these "messy" features of actual scientific practices we need a diagnosis of their sources.

3. Diagnosis: Reference, Meaning and Content in Real (and Messy) Life

I will argue below that what is going on in these examples of conceptual, semantic and epistemic messiness is a matter of commonplace partial failure of *accommodation* between concepts/language and relevant causal structures. This approach

rests on an *accommodationist* conception of reference and of natural kinds (Boyd, 1999a, 1999b, 2010) whose key features I describe below.

3.1 The key accommodationist idea and a non-human example

The basic idea of the accommodationist approach is that successful inductive/ explanatory practice requires *accommodation* between linguistic, conceptual, taxonomic resources and relevant causal factors and that the correct theory of reference and of natural kinds is whatever theory best explains how accommodation comes to be achieved. My accommodationist account entails the following.

1) Natural kinds, relations, etc., are components of successful accommodations between language use and causal structures in the relevant subject matter(s).
2) Referential hypotheses are hypotheses about how language-world relations contribute to accommodation in particular disciplinary matrices, so that
3) the naturalness of a kind is disciplinary-matrix specific.
4) The kind *natural kind* is itself a natural kind in the theory of how accommodation is achieved and the relation *reference* is similarly a natural relation in the same theory. Indeed,
5) natural kinds and reference are aspects of the very same phenomenon of accommodation.
6) The reference relation between natural kind terms and natural kinds is manifested in a dialectical relation between language use and features of the world such that the existence of a determinate referential relation between a term and a single natural kind is a somewhat rare special case.

It follows from such an approach that, in an important sense, natural kinds are social constructions, but this fact in no way diminishes their ontological status or supports an anti-realist treatment.

The basic accommodationist idea can perhaps best be appreciated by considering a special case concerning the semantics of signals in non-human animals. Paul Sherman and his associates (Sherman, 1977, 1985) have identified two sorts of alarm calls in Belding's ground squirrels. One, call it *a*, warns of aerial predators; the other, call it *t*, warns of terrestrial predators. Whence the semantics for these calls? Are calls *a* perfectly correlated with the presence of aerial predators and *t* with terrestrial predators? No, there are false positives and false negatives in each case. Instead what establishes the semantic relations are the facts that

1) *a* is frequently produced in response to aerial predators, not terrestrial ones.
2) *t* is frequently produced in response to aerial predators, not terrestrial ones.
3) Belding's ground squirrels exhibit different evasive responses to *a* and to *t*.
4) (1)–(3) explain how *S. beldingi* avoid predation.

What to note:

1)   The referential hypotheses explain achievements by indicating
2)   *accommodation* of signals, behaviours to relevant causal structures.
3)   The correct semantics is "discipline" specific. The causes and effects of alarm
     calls that are semantically relevant are just those that help to explain success
     in the "discipline" of predator avoidance.
4)   The relevant categories ("aerial predator," "terrestrial predator") and achieve-
     ments (both sorts of predator avoidance) are species-and-"discipline" spe-
     cific; "aerial predators" are aerial predators on ground squirrels not, for
     example, on insects.
5)   Still, the species-and-"discipline" specific nature of these categories and
     achievements does not diminish their ontological standing. The squirrels are
     real, and so are the relevant predators and patterns of predation and predator
     avoidance. One can be a realist about, e.g., the category "aerial predator on
     Belding's ground squirrels" without worrying that it is somehow "irreducibly
     rodential" and thus ontologically deflated.

It remains to see how the accommodationist account works in the case of human
language.

3.2 The accommodationist theory: initial simplified approximation

Here is the fundamental question: "How do classificatory practices and their lin-
guistic manifestations help to underwrite the reliability of scientific (and everyday)
inductive/explanatory practices?" When we inquire about the definition of a natural
kind, K, we are asking something like this: What commonalities in the causal
profiles of things we classify as Ks explain such inductive and explanatory suc-
cesses as we have achieved using the term K? H_2O is the definition of the kind water
because (1) to a good first approximation we tend to classify substances under the
term "water" (or related term in other languages) just in case they are mainly H_2O
and (2) this fact helps to explain our inductive/explanatory successes with respect
to the term "water."

   Of course the definition of a natural kind, K, depends on the actual inferential
practices of the relevant scientific communities: on the inferential architecture of
the relevant discipline. So the definition of any given K depends on the character-
istic inferential connections between the term referring to K and all of the other
natural kind terms within the discipline. The correct referential semantics for
discourse within a discipline will, to a simplified first approximation, be an assign-
ment, to each natural kind term, of a family of properties such that (1) the actual
usage of each term approximately "tracks" the family assigned to it, and (2) the fact
that this pattern of tracking occurs explains the reliability – such as it is – of the
discipline's inferential practices. Here is a more precise way of saying all this.

Let M be a disciplinary matrix and let $t_1, \ldots, t_n$ be the natural kind terms deployed within the discourse central to the inductive/explanatory successes of M. Then the families $F_1, \ldots, F_n$ of properties provide definitions of the kinds referred to by $t_1, \ldots, t_n$, and determine their extensions, just in case:

1. (Epistemic access condition) There is a systematic, causally sustained, tendency – established by the causal relations between practices in M and causal structures in the world – for what is predicated of t_i within the practice of M to be approximately true of things which satisfy F_i, $i = 1, \ldots, n$. In particular, there is a systematic tendency for things of which t_i is predicated to have (some or most of) the properties in F_i.
2. (Accommodation condition) This fact, together with the causal powers of things satisfying these explanatory definitions, causally explains how the use of $t_1, \ldots, t_n$ in M contributes to accommodation of the inferential practices of M to relevant causal structures. It explains whatever tendency there is for participants in M to identify causally sustained generalizations, to obtain correct explanations, or to obtain successful solutions to practical problems.

A note regarding "reality": Just as in the case of alarm calls for Belding's ground squirrels, the disciplinary-matrix specificity of natural kinds does not diminish their ontological standing. We are real, and so are our projects and linguistic usages and the things we study, so natural kinds are real too. Fair play for humans!

3.3 How things get messy

Imagine a disciplinary matrix, M_{simple}, in which things work exactly as the simplified accommodationist conception has it. The inferential practices in M_{simple} are sufficiently uniform and sufficiently reliable that it is pretty nearly determinate which relations satisfy the epistemic access condition for M_{simple}, and it is pretty nearly determinate how those practices contribute to epistemic successes in M_{simple}. Each term use in M_{simple}, corresponds (via conditions 1 and 2) to a single family of properties. In particular, no term fails altogether to refer, nor does any term participate in partial denotation (*sensu* Field, 1973), nor does any term have whatever sort of messy semantics currently infects the term "gene" (which might be characterized as lying between denotation failure and wholesale partial denotation). Suppose further that the conceptual meanings of terms and sentences in M_{simple} are coherent and pretty nearly true or reliable. In that case for a competent practitioner in M_{simple} who accepts an M_{simple} sentence, or has a belief which she knows how to express with an M_{simple} sentence, there would be no doubt both that the sentence or belief has determinate content and truth

conditions and that she knows the meaning or content of the sentence or of her belief.[2]

For all of the reasons presented in section 1 we can see that such an M_{simple}, were there to be one, would be profoundly atypical among scientific disciplinary matrices. Things are in fact always much messier. And yet (!) often things get (epistemically and semantically) better. Sometimes, quite often in fact, we start out with scientific terms hanging very loosely on the world (some not at all) and with concepts and inferential practices some of which are profoundly wrong, even incoherent, and we still make progress. We sometimes resolve partial denotations, discard non-referring terms, introduce a suite of new terms and concepts to "clean up" earlier unclarity, uncover and correct some malignant meanings, etc. We can only do that if there is already some sort of epistemically favourable *accommodation* between our linguistic, conceptual and inferential practices and some aspects of the world, but that initial accommodation need not be neat at all. How then should we think of issues of reference or of conceptual understanding?

3.4 Reference, understanding and rationality as temporally extended dialectical processes

What the accommodation thesis indicates is that reference (for natural kind terms) is the causal, explanatory correspondence relation between language use and the world which explains how the accommodation of language to relevant causal structures is achieved. I have argued elsewhere (Boyd, 1993, 2001b, 2010, 2012) that reference should thus be seen as the dialectical *process* of accommodation between the use of such terms and causal structures which is achieved by reliable inductive and explanatory practice. Both partial denotation *and denotational refinement and other strategies for "cleaning up" language and concepts* are aspects of the ongoing process of reference. The achievement of a referential situation in which a natural kind term enjoys a determinate definition and referent is thus a *(very) special case* of the phenomenon of reference.

Exactly the same considerations hold with methodological aspects of rationally understanding scientific sentences (and corresponding beliefs). Understanding of a scientific statement involves not somehow apprehending its content and truth-conditions (even if these are determinate), but instead involves understanding something about the rationally appropriate inferential relations between it and other sentences or beliefs. As cases of malignant meanings indicate, fully rational understanding of this sort may be very poorly approximated indeed within a disciplinary matrix for quite some time (think again of confusions about nature and nurture). So

---

2   I am simplifying somewhat here because in cases where one or more of the relevant terms refers to a homeostatic property cluster (HPC) kind, property or relation (see Boyd, 1999a, 2001b, 2010) there will be failures of bi-valence despite the cleanness of the relevant semantic relations.

we should think of scientific understanding and scientific rationality with respect to inferential relations not, in the first instance, as states which practices in a disciplinary matrix at a time either (almost) perfectly possess or lack altogether. Instead, understanding and rationality are temporally extended dialectical processes like reference where, again, really clean and neat achievements are relatively rare.

3.5 The social and environmental dialectics of reference understanding and rationality

Scientific research (like research in any other discipline) is, of course, socially orchestrated and environmentally embedded. The dialectical improvements with respect to reference, understanding and rationality just discussed – like all cases of scientific progress – reflect the coordinated work of lots of research communities, each with its own particular inferential norms, and each engaged in epistemically relevant interactions with the relevant subject matter(s). So, for example, if the semantics of terms like "gene" ever gets sorted out (probably by the introduction of much richer and better accommodated terms and concepts) the work will be done by lots of different research communities all studying various genetic phenomena.

3.6 The dialectics of projectibility judgments: a crucial special case

I have emphasized that the sorting out of partial denotation and other referential oddities and of malignant meanings rests on ongoing causal interactions between research communities themselves and between those communities and their subject matter(s) – interactions of just the sort which are taken, by fans of wide content, to be individuative of propositional attitudes. A further exploration of the dialectics involved will be important for issues later on in the present article.

The first thing to notice is that the resolution of the semantic and methodological issues in question is simply a special case of the ways in which theoretical and methodological issues get resolved in scientific practice. The sorting out of the semantics of "gene" will be simply a special case of theoretical advances in genetics, not of some distinctly semantic inquiry. So the lesson about social and environmental dialectics extends to all cases of theoretical or methodological progress in the sciences. Let us see how that goes and what it implies about how scientists can abide by (and change) community norms.

3.6.1 Projectibility and evidence, initial approximation. To a pretty good first approximation a theory, T, counts as confirmed by evidence, E, given the evidential standards of a scientific community at a time just in case:

1)   T is projectible by that community's standards, and
2)   E favours T over all relevant alternatives (= rival theories also projectible by those standards), and

3)   E was gathered so as to control for artifacts suggested by theories projectible by those standards.

3.6.2 Projectibility and truth. What has projectibility to do with truth? Here is another way of putting these standards evaluating evidence.

Basic methodological rule of science: Carefully choose from among relevant alternatives (= theories recommended by best current theories), controlling for effects suggested by best current theories.

Why is this rule reliable? In particular, what makes relying on current theories and inferential practices in this way reliable ? Answer: Surely *not*: their currency. Instead to see why (and when) these methods are reliable we need yet more rules.

Basic "falsificationist" rule for objective testing: Try to falsify T. Test T where it is most vulnerable, i.e., under circumstances where it is most likely to go wrong if it is not (relevantly, approximately) true.

Rule for identifying such circumstances: Identify the most plausible alternatives to T: its projectible rivals. Similarly, for identifying experimental artifacts.

When is this practice epistemically reliable? To a pretty good first approximation (better ones later) the basic methodological rules is reliable just when the prevailing theories are accurate enough and the prevailing inferential practices are reliable enough that (a) pretty often an approximately correct answer to a scientific question will be among the projectible answers to it, and (b) pretty often one can avoid experimental artifacts by controlling for those suggested by projectible theories.

3.6.3 Three social and dialectical complexities. Three respects in which this picture is incomplete will be important later in this article.

Multiple communities: Often, a researcher will need to participate in the inferential practices of more than one research community in order to address an important scientific question. I do not merely mean that, e.g., a geologist might need to apply some standard results or practices in chemistry or physics. I mean that she may need to engage seriously with specialists in some other discipline or subdiscipline and/or with its literature, where inferential practices, theoretical conceptions and conceptual meanings may differ (perhaps quite seriously) between her (sub)discipline and the other. Of course this sort of engagement would be required in cases in which systematic errors like malignant meanings in one (sub)discipline were corrected by critiques from other perspective of other (sub)disciplines, but it is quite commonplace in everyday scientific practice for there to be substantial differences in practices and meanings and norms between disciplines even when no malignancy is involved. So following the disciplinary norms of one scientific

community will often require that one engage with the disciplinary norms of other communities.

Diachronic variability of methodological norms: There is a dialectical relationship between theories, methods and their subject matter(s) within a discipline so that – when things go well – there are continuing theoretical and methodological advances. So, even for a practitioner who is engaged seriously with only a single discipline or subdiscipline, adherence to internal disciplinary norms requires continuing immersion in the discipline's practices and literatures. This is not true just of the "ideal" researcher: it is true of the ways in which ordinary researchers exhibit ordinary levels of scientific rationality.

Dialectical and social character of synchronic norm applications: Even the synchronic application of disciplinary norms, even within a single community, crucially involves interactions between community members. It is normative that theories should be assessed for projectibility and tested against projectible alternatives. Identifying such alternatives is a matter of theory invention. What adherence to this norm requires of the individual researcher is not just that she evaluate proposed theories against those projectible alternatives that she invents. Instead disciplinary standards, both formal (like referees' reports on submitted papers or evaluations of grant proposals) and informal (like the standards that prevail in research group meetings, academic conferences, ordinary conversations with colleagues, etc.), require that one take into account relevant alternatives suggested by others. Again, this is true not just for "ideal" researchers: it is true of the ways in which ordinary researchers exhibit ordinary levels of scientific rationality. Implementing individual scientific rationality is an essentially social activity.

4. Interlude: Where We Are So Far

4.1 Linguistic messiness
Reference failure, partial denotation and other semantic anomalies are widespread. So, too, is limited or defective understanding of truth conditions and/or of appropriateness of inferential practices even when such anomalies are absent. Community conceptual and inferential norms are not justifiable *a priori*. Sometimes they are profoundly incoherent so that reference is achieved *despite* local community norms. Almost always in any real case, scientists will participate to some extent in different research communities with different norms.

4.2 Propositional attitude messiness
Exactly corresponding messiness attaches to propositional attitudes and to our knowledge of the content or truth conditions of our own propositional attitudes.

At least where the subject matter is esoteric our propositional attitudes often do not have determinate contents or truth conditions for us to know and pretty often our understanding of such contents, even when they are (or are almost) determinate, is profoundly compromised by malignant conceptual meanings. So, a correct theory of our knowledge of our own propositional attitudes should predict that pretty often we barely know what we think.

4.3 Makings things better

In so far as we do have some (typically partial) grasp on the contents and truth conditions for our beliefs, conjectures, etc., and for the sentences which express them, it is because we are embedded in discourse communities with some epistemic access to the phenomena we are thinking or talking (or writing) about. In so far as we can expect to improve our self-understanding in this regard it is because of our continued embedding in such research communities.

4.4 Knowledge and the individuation of propositional attitudes

So far we have seen four things.

1) The case for individuating propositional attitudes widely rests on *a posteriori* claims about how accommodation of linguistic and conceptual items to relevant causal structures is achieved. No mention of "philosophical intuitions" about Twin Earth cases is necessary.
2) Anyway, such intuitions grounded in everyday norms of language use would not be a source of *a priori* knowledge even if they got it right about important cases. That is because
3) community semantic and inferential norms do not reflect *a priori* knowledge even when they get things right. Anyway, they sometimes are profoundly mistaken.
4) Finally, individuating propositional attitudes widely does not raise any problems about our knowledge of our own propositional attitudes. The features of social and environmental embedding reflected in wide criteria for individuation are just those features such that
    a) our involvement with them makes possible such partial knowledge as we do have and
    b) continuing involvement with them underwrites whatever chances we have of improving that knowledge.

It remains to be seen whether wide individuation is appropriate for the scientific study of propositional attitudes.

5. Content Externalism as a Naturalistic Causal-Explanatory Psychological Hypothesis

5.1 Basic strategy

As I have said, this part of the present article is speculative. I will try here to articulate what I believe is the best scientific case for wide individuation of propositional attitudes. Here is an outline of the basic arguments.

1) *Basic cognitive science assumption:* Psychological explanations in terms of propositional attitude contents are causal explanations.

2) *Explanatory scope:* The aim of cognitive science with respect to propositional attitudes is not just to explain and predict behaviours and choices but also to explain the extent of, and limitations to, the epistemic successes of human rationality. In particular we should expect propositional attitude explanations of systematically successful rational problem solving in theoretical and practical disciplines as well as explanations of those cases where rational problem solving fails to achieve successes.

3) *Rationalizing explanations:* So, in so far as psychological explanations cite features of individual human rationality the rationality in question must be of the sort that people actually exhibit. Scientifically appropriate rationalizing explanations are causal explanations reflecting the causal roles which propositional attitudes actually play in human rationality, such as it is.

4) *Basic accommodationism re individuation of propositional attitudes* (= special case of the accommodationist conception of natural kind terms): Propositional attitudes should be individuated in terms of whatever factors sustain the causal profiles that explain how those attitudes contribute to determining behaviours, choices, *and* rational reasoning *and* epistemic successes and failures.

5) *Wide individuation:* Those factors include the features of social embedding, community norms, and causal interaction with subject matter(s) posited by wide-scope conceptions of individuation.

In the present short article I will simply assume that 1–4 are true and argue for 5. Of course, if scientific approaches to psychological explanation of the sort anticipated in current cognitive science are unworkable, or if ordinary rationalizing explanations cannot be incorporated into cognitive science causal explanations, then the arguments I am going to offer will fail, but that is a topic for another article.

5.2 Modelling propositional attitude transitions via community discourse structure

So here is the case for the causal importance of wide-individuative factors in psychological explanations. In the first place I suggest (and anyone who thinks of propositional attitude explanations as, in some sense or other, rationalizing explanations should agree) that a crucial feature of the way we rely on propositional attitude attributions in order to explain (or predict for that matter) how people reason, choose, behave, get things right or wrong, etc., involves what I will call *community discourse modelling*. Here is a somewhat simplified picture of how community discourse modelling works to explain how a person, P, made some inferences, or reached some decision.

1) We attribute to P some propositional attitudes, $a_1, a_2, a_3, \ldots, a_N$, which we characterize by sentences, $S_1, S_2, S_3, \ldots, S_N$, such that each sentence, S_i, would be a good candidate for expressing the content of the attitude a_i, in some contextually definite community(ies). (Think: "She believed S_1 and S_2 but she doubted S_3 and conjectured that S_4 might be true instead.")

2) We explain aspects of P's inferences or practical reasoning by assuming that she has a causally important tendency, given the attitudes, $a_1, a_2, a_3, \ldots, a_N$, to make inferences or reach practical conclusions that accord closely with those that would be licensed by the inferential norms in the relevant community(ies) given, as premises, the sentences $S_1, S_2, S_3, \ldots, S_N$. Roughly, we model P's reasoning processes by analogy to discourse norms in the community(ies) in question.

5.3 Why should we do this?

Suppose for the sake of argument that this sort of rationalizing explanation is appropriate. Whence its appropriateness? There are three plausible explanations.

1) Rationalizing explanations are not causal.

2) These explanations are causal and they work because individual rationality is, at least to a sufficiently good first approximation, the internal deployment of acquired community norms. We assume that P has internal states $a_1, a_2, a_3, \ldots, a_N$, whose internal representations are somehow relevantly computationally related *internally* to her capacity to deploy sentences like $S_1, S_2, S_3, \ldots, S_N$ and that her internal information processing of $a_1, a_2, a_3, \ldots, a_N$ conforms to the community norms for $S_1, S_2, S_3, \ldots, S_N$ because she has *internalized* those norms as computational inference rules for $a_1, a_2, a_3, \ldots, a_N$.

3) These explanations are causal and they work because over time individuals do tend to reason in accordance with communities' norms. Sometimes, the internal mechanisms posited by answer 2 do operate. *But*, in general,

substantial actual participation in such communities is absolutely crucial in realizing rational structure of an *individual's* attitude transitions, inferences, judgements, etc. *So are* (sometimes over longer time intervals) epistemically relevant causal interactions (almost always community-mediated) with the relevant subject matter(s).

Explanation 1 I have proposed to set aside for the purposes of the present article.

Explanation 2 would seem to rationalize a narrow-content approach to scientific psychological explanations: our discourse-modelling explanations *do* tacitly refer to community norms and practices (via the use of sentences $S_1, S_2, S_3, \ldots, S_N$) but only as a way to refer to the *purely internally* realized and individuated propositional attitudes and to *purely internally* realized and individuated causal relations between them. According to Explanation 2, we successfully deploy wide-content features, not to individuate propositional attitudes, but to make possible convenient reference to the purely internal factors that properly individuate propositional attitudes.

Explanation 3 is the one which would underwrite individuation for propositional attitudes in scientific psychological explanations.

5.4 For wide individuation

OK, why is Explanation 3 better? As the points about multiple communities, the diachronic variability of methodological norms, and the dialectical and social character of synchronic norm applications in section 2.5.2 indicate, the implementation of individual inductive and explanatory rationality depends crucially on the embedding of individual practitioners in their community(ies) and their community-mediated epistemic engagement with relevant subject matters. Neither the outcome of individual rational reasoning, nor the epistemic achievements often thereby obtained, are scientifically explicable without positing exactly the sort of social and environmental embedding indicated by wide-content conceptions.

It is worth elaborating this point with respect to two aspects of rationalizing explanations. Remember that we are assuming here that those explanations are causal, so that the crucial question about the individuation of propositional attitudes is this: what properties of human propositional attitudes underwrite those of their causal powers that are reflected in successful rationalizing explanations? Successful rationalizing explanations *presuppose* (correctly) that people are often guided by the relevant rational norms and they *presuppose* (also correctly – let us hope, because otherwise cognitive science and the rest of the sciences are epistemically irrelevant) that, pretty often, their adherence to those norms results in practical or theoretical epistemic achievements.

These two central presuppositions are closely related, of course. In one direction the relationship is obvious: if people could not implement rational norms then the implementation of those norms could not explain practical and theoretical successes. The connection in the other direction is this. In many cases community norms are sensitive to theoretical and practical successes and failures: norms get revised, or rejected, or (on the other hand) enhanced in their importance, depending on apparent failures or successes resulting from their implementation. Thus the variability over time of a community's norms will sometimes depend on the epistemic successes or failures which they underwrite. Similarly the dialectical interaction between the (sometimes competing, sometimes complementary) norms of different communities will often depend on the respects in which they apparently result in epistemic successes or failures and this will, in turn, often depend on the extent of their actual successes or failures.

So the crucial question about propositional attitude individuation is this: what properties of propositional attitudes underwrite the truth of the two presuppositions in question? What makes it possible for humans to apply the relevant community norms of rationality? What explains the (certainly imperfect) tendency for the application of such norms to contribute to theoretical and practical successes?

The key to preferring Explanation 3 to Explanation 2 with respect to the first question is three-fold. In the first place, it is not possible for a person to learn all the subtleties of even one community's norms and to apply them effectively in the absence of social interactions within her own community. In so far as her community's norms change in response to interactions with the relevant subject matters, her ability to follow those norms will depend on her continuing involvement with that community. Finally, and crucially, in actual practice scientists need to be responsive to the changing norms and the achievements of several different communities. That can only be achieved by engagement (not necessarily always deferential) with the norms and the achievements of communities pretty distant from one's own where there is no prospect of internalizing all the relevant norms and achievements.

With respect to the second question, no explanation of such success-making featured as the norms of scientific (or other) communities can exhibit can be given without some externalist elements: some systematic causal connection between community practices and the relevant subject matter must be posited.

It is important to see one way in which the answers to these questions are related. The capacity of scientists to follow a community's norms will depend, of course, on those norms exhibiting some substantial stability over time so psychological explanations of the sort we are considering presuppose some important underlying stability in norms. The practice of revising a community's norms in response to judgments about their records of success-making (or failure-making) of course reduces that stability, so rationalizing explanations presuppose

that such revisions are usually modest. Why is that so? Well, often enough, it is because gradual improvement in background theories and methods (via previous revisions) has rendered the norms in question pretty reliable epistemically, so that dramatic revisions are less often called for. Of course that is by no means the only explanation (consider methodological norm stabilities in various different domains of dogmatic theology for example, or the role of social ideology in stabilizing methodological norms in human genetics through much (all?) of the twentieth century), but the role of epistemically relevant contact with the subject matter(s) can often be important, just as a externalist perspective would suggest.

So the embedding of researchers in communities of the sort singled out by fans of wide individuation is explanatorily central in explaining how researchers can follow the communities' norms. In so far as we are concerned to explain the stability of those norms over time and (especially) to explain how human rationality can contribute to long-term scientific successes, the full dialectical complexities of diachronic interactions of scientific communities with each other and their subject matters will come into play. In all cases the explanatory centrality of social and environmental embedding is sustained.

5.5 Wide individuation and the scope of the "laws of thought"

Here is another way of thinking about the individuation of propositional attitudes. If the rational discourse modelling thesis is approximately right, then propositional attitude explanations of beliefs, desires, choices *and of their (favourable or unfavourable) epistemic properties* posit something like "laws of thought" corresponding to community discourse norms. Of course these "laws" would not be laws of the sort anticipated by logical empiricist reconstructions of science. Instead, propositional attitude explanations would posit persisting *tendencies* of subjects to have their information processing states conform in the relevant ways to the discourse norms of the relevant communities.

The term "tendencies" here is important for two reasons. In the first place, of course, our ordinary practices of propositional attitude explanations, and the corresponding practices of cognitive scientists, do not posit that the only factors involved in individual deliberations, choices, belief formations, etc., are those reflecting the deliberative norms of the subject's communities and the interactions of those communities with their subject matters. Affective states, cognitive limitations (or special cognitive gifts), personal commitments and similar factors are also posited as possible sources of departures from community norms. To some extent the psychology of science literature on "confirmation bias" may be thought of as involving the study of such factors (see Koslowski, 2012, 2013; Koslowski and Thompson, 2002; Koslowski and Masnick, 2010 for an overview). So, in one sense, the scope of rational discourse modelling explanations is limited: they are supposed

to account for some, but not all, of the important factors accounting for individuals' choices, beliefs, inferences, decisions, etc.

On the other hand, I have been assuming here that, in another sense, the explanatory scope of propositional attitude explanations is pretty broad. I have been assuming that when we (qua everyday inquirers or qua cognitive scientists) apply rational discourse modelling to (help to) explain beliefs, inference patterns, and epistemic successes and failures, etc., we intend to explain not just such features of simple everyday reasoning but also the same features of temporally extended episodes of reasoning regarding complex theoretical or practical issues. Here again, the literature on the psychology of science is relevant: it surely seems that cognitive scientists aim to explain the psychology of actual scientific reasoning in part by explaining the features of deliberative reasoning which contribute to epistemic successes and failures in practice (again see Koslowski, 2012, 2013; Koslowski and Thompson, 2002; Koslowski and Masnick, 2010).

Recognizing this breadth of explanatory scope is important to assessing the arguments presented here for wide individuation of propositional attitudes. Consider again the internalist second option explored in section 4.2 for explaining the appropriateness of rational discourse modelling explanations: that they are causal explanations and that they work because individual rationality is, at least to a sufficiently good first approximation, the internal deployment of acquired community norms. What I claim here is not that there are no cases in which both the nature of a subject's judgments, choices, inferences, etc., and their epistemic reliability are explicable just by her deployment of internalized community norms. What I claim is that in many centrally important cases – including cases of deliberative reasoning in the sciences and other intellectual disciplines – the causal powers of an individual's propositional attitudes presupposed by rational discourse modelling explanations *are* possessed by such attitudes, *but* only because of the relevant features of her social and environmental embedding.

6. Issues of Time Scale

Here is a rebuttal to what I just said: "When I'm working on a serious scientific or philosophical problem I can exhibit the sort of rationality in inferences and judgments anticipated in rational discourse modelling even when I am alone in my study interacting neither with the relevant discourse community(ies) nor with the phenomena about which I am reasoning. So exhibiting the causal powers anticipated in rational discourse modelling does not require the posited sorts of social or environmental embedding." I have already addressed this challenge by agreeing that the sorts of factors posited by the internalist option in 4.2 often operate to sustain (epistemically effective) rational compliance with community norms over

sufficiently short periods of time, but it is probably valuable to explore further just how, when working alone in your study, you can still be related to external social and environmental factors of the sort revered by externalists about propositional attitudes.

First, actual ongoing participation in an inferential community need not always involve being in close proximity with others in the community. Reading journals, publishing in them, corresponding by (e- or snail) mail count. The point is, after all, keeping up with and conforming to (most of) the inferential norms in question, not (necessarily) chatting. Second, the contact with the relevant subject matter(s) studied by the community(ies) in question need not be something that all members participate in. Experimentalists and theoreticians need not switch jobs from time to time in order to constitute a research community.

Nevertheless, you might conceivably live for a year as a hermit – reading nothing, corresponding and talking with no one – while still working out a fully rational contribution to some community's research literature while violating no community norms except perhaps norms of politeness. But – the externalist (correctly!) maintains – cases like that (where internalized norms suffice to under-write long-term rationality with respect to a particular project) fail to reflect the key general point about individual human rationality: that individual rationality is essentially social!

Acknowledgements

References

ALCOCK, J. (2001) *The Triumph of Sociobiology*. Oxford: Oxford University Press.
BOYD, R. (1993) "Metaphor and Theory Change" (second version). In A. Ortony (ed.), *Metaphor and Thought*, 2nd edn (pp. 481–532). New York: Cambridge University Press.
BOYD, R. (1999a) "Homeostasis, Species, and Higher Taxa." In R. Wilson (ed.), *Species: New Interdisciplinary Essays* (pp. 141–185). Cambridge, MA: MIT Press.
BOYD, R. (1999b) "Kinds, Complexity and Multiple Realization: Comments on Millikan's *Historical Kinds and the Special Science*s." *Philosophical Studies* 95(1–2): 67–98.
BOYD, R. (2001a) "Reference, (In)commensurability, and Meanings: Some (Perhaps) Unan-ticipated Complexities." In P. Hoyningen-Huene and H. Sankey (eds), *Incommensurability and Related Matters* (pp. 1–63). Dordrecht: Kluwer.
BOYD, R. (2001b) "Truth through Thick and Thin." In R. Schantz (ed.), *What Is Truth?* (pp. 38–59). Berlin and New York: Walter de Gruyter.

BOYD, R. (2010) "Realism, Natural Kinds and Philosophical Methods." In H. Beebee and N. Sabbarton-Leary (eds), *The Semantics and Metaphysics of Natural Kinds* (pp. 212–234). London: Routledge.

BOYD, R. (2012) "What of Pragmatism with the World Here." In M. Baghramian (ed.), *Reading Putnam* (pp. 39–94). London: Routledge.

BROWN, J. (2004) *Anti-Individualism and Knowledge*. Cambridge, MA: MIT Press.

BULLER, D. E. (2005) *Adapting Minds: Evolutionary Psychology and the Persisting Quest for Human Nature*. Cambridge, MA: MIT Press.

BURGE, T. (1979) "Individualism and the Mental." *Midwest Studies* IV: 73–121.

BURGE, T. (1986) "Intellectual Norms and the Foundations of Mind." *Journal of Philosophy* 83: 697–720.

BURGE, T. (1988) "Individualism and Self-Knowledge." *Journal of Philosophy* 85: 649–663.

BURGE, T. (1996) "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96: 91–116.

EGAN, F. (1995) "Computation and Content." *Philosophical Review* 104: 181–203.

FIELD, H. (1973) "Theory Change and the Indeterminacy of Reference." *Journal of Philosophy* 70: 462–481.

GIBBONS, J. (2001) "Externalism and Knowledge of the Attitudes." *Philosophical Quarterly* 51: 13–28.

KITCHER, P. (1985) *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.

KOSLOWSKI, B. (2012) "How Explanation Makes Information Evidentially Relevant." In R. Proctor and E. J. Capaldi (eds), *The Psychology of Science: Implicit and Explicit Reasoning* (pp. 175–204)*.* Oxford: Oxford University Press.

KOSLOWSKI, B. (2013) "The Intersection of Formal Principles and Non-Formal Background Knowledge in Scientific Reasoning." In G. Feist and M. Gorman (eds), *Handbook of the Psychology of Science and Technology* (pp. 151–192). New York: Springer.

KOSLOWSKI, B. and MASNICK, A. (2010) "Causal Reasoning." In U. Goswami (ed.), *Handbook of Child Cognitive Development*, 2nd edn (pp. 257–281). Oxford: Blackwell.

KOSLOWSKI, B. and THOMPSON, S. (2002) "Theorizing Is Important, and Collateral Information Constrains How Well It Is Done." In P. Carruthers, S. Stich and M. Siegal (eds), *The Cognitive Bases of Science: Interdisciplinary Approach* (pp. 171–192). Cambridge: Cambridge University Press.

KRIPKE, S. (1980) *Naming and Necessity*. Oxford: Blackwell.

MCLAUGHLIN, B. and TYE, M. (1998) "Externalism, Twin Earth, and Self-Knowledge." In C. Wright, B. C. Smith and C. Macdonald (eds), *Knowing Our Own Minds (Mind Association Occasional)*, pp. 285–320. Oxford: Oxford University Press.

NUCCETELLI, S. (ed.) (2003) *New Essays in Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press.

PUTNAM, H. (1962) "The Analytic and the Synthetic." In H. Feigl and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science*, Vol. III (pp. 358–397). Minneapolis: University of Minnesota Press.

PUTNAM, H. (1975) "The Meaning of 'Meaning'." In K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, Vol. 2 (pp. 215–271). Minneapolis: University of Minnesota Press.

RICHARDSON, R. C. (2007) *Evolutionary Psychology as Maladapted Psychology*. Cambridge, MA: MIT Press.

SAWYER, S. (2002) "In Defence of Burge's Thesis." *Philosophical Studies* 107: 109–128.

SEGAL, G. (2003) "Conceptual Errors and Social Externalism." *Philosophical Quarterly* 53: 265–273.

SHERMAN, P. (1977) "Nepotism and Evolution of Alarm Calls." *Science*, 197: 4310.

SHERMAN, P. (1985) "Alarm Calls of Belding's Ground Squirrels to Aerial Predators: Nepotism or Self-Preservation?" *Behavioral Ecology and Sociobiology* 17:4.

WIKFORSS, Å. M. (2001) "Social Externalism and Conceptual Errors." *Philosophical Quarterly* 51: 217–231.

WIKFORSS, Å. M. (2004) "Externalism and Incomplete Understanding." *Philosophical Quarterly* 54: 287–294.

WRIGHT, C., SMITH, B. C. and MACDONALD, C. (1998) *Knowing Our Own Minds (Mind Association Occasional)*. Oxford: Oxford University Press.