Metadata of the chapter that will be visualized online

Chapter Title	Stirring the Possum: Responses to the Bianchi Papers	
Copyright Year	2020	
Copyright Holder	Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Devitt**
	Particle	
	Given Name	**Michael**
	Suffix	
	Division	Philosophy Program
	Organization/University	City University of New York Graduate Center
	Address	New York, NY, USA
	Email	MDevitt@gc.cuny.edu
Abstract	This paper is made up of responses to the papers in this volume. Part I is concerned with the philosophy of linguistics, particularly with the linguistic conception of grammars and the psychological reality of language. Part II is concerned with the theory of reference for proper names, including issues of reference borrowing, grounding, the qua-problem, and causal descriptivism. Part III is concerned with the theory of meaning, particularly the matters of direct reference, the contingent a priori, rigidity for kind terms, narrow meaning, and the use theory. Part IV is concerned with methodology, particularly with "putting metaphysics first", the role of intuitions, and the contribution of experimental semantics. Part IV tackles two issues in metaphysics: the definition of scientific realism and biological essentialism.	
Keywords (separated by " - ")	Philosophy of linguistics - Theory of reference - Theory of meaning - Intuitions - Experimental semantics - Scientific realism - Biological essentialism	

Chapter 19
Stirring the Possum: Responses to the Bianchi Papers

Michael Devitt

Abstract This paper is made up of responses to the papers in this volume. Part I is concerned with the philosophy of linguistics, particularly with the linguistic conception of grammars and the psychological reality of language. Part II is concerned with the theory of reference for proper names, including issues of reference borrowing, grounding, the qua-problem, and causal descriptivism. Part III is concerned with the theory of meaning, particularly the matters of direct reference, the contingent a priori, rigidity for kind terms, narrow meaning, and the use theory. Part IV is concerned with methodology, particularly with "putting metaphysics first", the role of intuitions, and the contribution of experimental semantics. Part IV tackles two issues in metaphysics: the definition of scientific realism and biological essentialism.

Keywords Philosophy of linguistics · Theory of reference · Theory of meaning · Intuitions · Experimental semantics · Scientific realism · Biological essentialism

Thank you very much indeed to my friends who contributed such rich and thought-provoking papers to this volume. And a special thanks to Andrea Bianchi for conceiving of this volume and pulling it off. He is a wonderful editor![1]

The papers are *so* thought-provoking that, I confess, I found responding to them rather overwhelming. They obviously raise many more issues than I can deal with here. I focus my responses on areas where I think I have something new to say or

[1] So wonderful that I forgive him for rejecting "Stirring the Possum" as the title of this volume (although he did let me use it for my responses). It's an Australian expression and he was worried that its meaning would be lost on non-Australians. But that has not been my experience. And its meaning is only a Google click away for those who can't figure it out. I am indebted to Bianchi also for some helpful comments on a draft of these responses.

M. Devitt (✉)
Philosophy Program, City University of New York Graduate Center, New York, NY, USA
e-mail: MDevitt@gc.cuny.edu

A. Bianchi (ed.), *Language and Reality from a Naturalistic Perspective*, Philosophical Studies Series 143, https://doi.org/10.1007/978-3-030-47641-0_19

where I detect misunderstandings. So my responses to papers I largely agree with will be very brief. And where I think – perhaps mistakenly! – that I have dealt adequately with a criticism elsewhere, I shall simply cite that discussion.

I shall organize my responses into sections according to topics. The authors of any of the volume's papers that I cite in a section are mentioned parenthetically in the section's title. Some authors are cited in more than one section.

19.1 Philosophy of Linguistics

19.1.1 The Linguistic Conception of Grammars (Collins, Rey)

19.1.1.1 Introduction

In my book, *Ignorance of Language* (2006a),[2] I urge seven "major conclusions" and entertain seven "tentative proposals" about language and linguistics. John Collins rich paper, "Invariance as the Mark of the Psychological Reality of Language", is the latest in our long-running exchange about some of these theses (Collins 2006, 2007, 2008a, b, Devitt 2006b, 2008a, b). He takes particular issue here with the two major conclusions that have turned out to be the most controversial. One of these is:

> *First major conclusion*: Linguistics is not part of psychology (40).

I urge that a grammar is about a nonpsychological realm of linguistic expressions, physical entities forming symbolic or representational systems (Chap. 2).[3] I later called this "the linguistic conception" of grammars. It stands in contrast to the received Chomskian view that a grammar is about a speaker's linguistic competence and hence about mental states. I later called this "the psychological conception".[4] The other conclusion that concerns Collins is:

> *Third major conclusion*: Speakers' linguistic intuitions do not reflect information supplied by the language faculty. They are immediate and fairly unreflective empirical central-processor responses to linguistic phenomena. They are not the main evidence for grammars. (120)

[2] All unidentified references to my work in Sect. 19.1 are to this book.

[3] See also Devitt 2003. Devitt and Sterelny 1989 is an earlier version of the argument, but contains many errors.

[4] I emphasize that

> the linguistic conception does *not* involve the absurd claim that psychological facts have nothing to do with linguistic facts. Some psychological facts *cause* linguistic facts (23–24), some "*respect*" them (25), some *partly constitute* them (39–40, 132–133, 155–157), some *provide evidence* for them (32–34), and some make them *theoretically interesting* (30, 134–135). But psychological facts are not the *subject matter* of grammars. The dispute is not over whether linguistics relates to psychology but over the way it does. (2008a: 207)

Part of Georges Rey's remarkably polemical, "Explanation First! The Priority of Scientific Over 'Commonsense' Metaphysics", is the latest in an even longer-running exchange (Rey 2006a, b, 2008, 2014, Forthcoming-a, Forthcoming-b, Devitt 2006a, 2008a, 2013a, 2014a, 2020). It has a few pages rejecting my first major conclusion and mentions his earlier rejection of the third.

I have already said a great deal about intuitions (2006a, c, d, 2008a, c, 2010a, b, 2012a, 2014a, b, 2015a, 2020) and so will say only a little more here (in Sect. 19.4.3). I shall discuss my first major conclusion, the linguistic conception of grammars, focusing on Collins' lengthy discussion but with passing attention to Rey's much briefer one. I discuss other aspects of Rey's paper in Sects. 19.4.1 and 19.4.2 below.

My linguistic conception has provoked a storm of criticism.[5] Yet, as Collins observes (note 2), I have complained that my critics, including Collins himself, have tended not to seriously address my *argument* for this provocative conclusion. Collins' paper in this volume is certainly not open to that complaint: he addresses my argument in great detail, for which I thank him. His paper is a novel and important contribution to the fundamental issue of what grammars are about.

19.1.1.2 The "Master Argument"

As Collins points out, (what he calls) my "master argument" (17–38) rests on the application of three general distinctions to humans and their language. The distinctions are, briefly: first, between the theory of a competence and the theory of its outputs or inputs; second, between the structure rules governing the outputs and the processing rules governing the exercise of the competence; third, between the "respecting" of structure rules by processing rules and the inclusion of structure rules among processing rules. ("Respecting" is a technical term here: processing rules and competence "respect" the structure rules in that they are apt to produce outputs governed by those structure rules; analogously, apt to process inputs so governed.) So, using Collins' numbering, premise (1) of my argument is that there are these three general distinctions. Collins accepts this but in Sect. 2.4 he resists premise (2): that these distinctions apply to humans and their languages. So he must resist the result of this application, his (3). And in Sect. 2.5 he resists my final step which he expresses as follows: (4), "a grammar is best interpreted as a theory of the structure rules of linguistic expressions, not of linguistic competence" (##10).

The basis for (2) is simple: the distinctions in (1) are quite general, applying to *any* competence; so they apply to linguistic competence. Why does Collins think not? (a) Collins doubts that the distinctions really are general. (i) They "appear not to apply to the mammalian visual system" which does not "produce external products" (##11). But they do apply to the vision system, nonetheless: that system is a

[5] Antony 2008; Collins 2006, 2007, 2008a, b; Longworth 2009; Ludlow 2009; Pietroski 2008; Rey 2006a, b, 2008; Slezak 2007, 2009; Smith 2006.

competence to process certain *inputs* – and, we should note, so partly is the language system – and the distinctions apply to these as much as to a competence to produce certain outputs (17). (ii) The distinctions also fail to apply to "organs" (##12). Perhaps, but the generalization is only about competencies and hence applies to an organ only insofar as it is a competence. (b) Collins thinks my application of these general distinctions to linguistics rests on dubious "analogical reasoning" (##12). True, I do argue for the distinctions using examples of competencies; my favorite is the waggle dance of the bees. But this is not aptly called "analogical reasoning". It is better seen as arguing for a generalization about *Fs* using examples of *Fs as evidence*, surely a sound scientific practice. (c) Collins claims "that the distinctions presuppose an externalism of structured outputs" (##15; see also ##13). But, strictly speaking, they do not. The distinctions apply even if a competence is never exercised. Thus there would still be a distinction between a theory of competence in a language and a theory of the structure rules of its linguistic outputs even if those outputs were only possible not actual.[6]

19.1.1.3 Linguistic Realism and Explanation

In any case, I think that Collins' deep objection to the linguistic conception lies elsewhere. Although the application of the three distinctions to human languages does not presuppose an external-to-the-mind linguistic reality, "linguistic realism", the linguistic conception as a whole does. For, according to that conception, grammars explain the nature of that linguistic reality and one can't explain what doesn't exist. Furthermore, it has to be the case that the *theoretical interest* of grammars comes primarily from such explanations. Collins argues that neither of these requirements, particularly the second, is met. My view is alleged to rest on the "misunderstanding … that there are linguistic phenomena as such, which a linguistic theory may target directly, with psychological phenomena being targeted only indirectly" (##2). Here is a strong statement of his view: "I reject linguistic externalia … simply because they are explanatorily otiose: they neither constitute phenomena to be explained nor explain any phenomena" (##10).

Rey's objection to the linguistic conception is similar. In his earlier-mentioned exchange with me, Rey has mounted the most sustained attack on linguistic realism that I know of; and I have responded. In the present paper, he again urges his antirealism. And he insists that grammars explain psychological phenomena: "Chomskyan linguistics is per force about psychology because that's simply where the relevant law-like regularities lie" (##12).

[6] It is apparent that the competence/performance distinction is central to my application of the three distinctions to linguistics. So it is odd that Collins often writes as if I have overlooked this much-loved distinction (##6, 12, 30–32).

The center of disagreement is over two related explanatory issues: What is worthy of explanation in linguistics? What do grammars actually explain? The fate of each conception, linguistic and psychological, largely hinges on these questions.

Starting in *Ignorance* (28–35, 134–135, 184–192), and continuing in various responses to critics (2006b, 2007, 2008a, b, c, 2009a, 2013a, b), I have argued for linguistic realism, for the view that grammars (partly) explain that reality, and for the view that it is primarily in virtue of this that grammars are theoretically interesting. I shall not repeat these arguments but the key ideas are as follows. First, we posit this linguistic reality to explain behavior, particularly communicative behavior: a noise (or inscription) is produced *because* it has certain linguistic properties, including syntactic ones, and it is *because* it has these properties, and hence is a linguistic expression (symbol), that an audience responds to that behavior as it does. Second, a grammar is a straightforward account of the syntactic properties of that expression and hence *a (partial) explanation of the nature of the expression in virtue of which it plays its striking causal role in human lives*.

Collins' guiding idea is that "the ontology of a theory is ultimately what is invariant over and essential to the explanations the theory affords" (##2). I have no quarrel with that. But I think that what are invariant and essential to the explanations that grammars afford are linguistic expressions, like the sentences, nouns, determiners, etc. on this very page. Collins introduces a notion:

> *Primitive explanation*: *T primitively explains D* iff *T* explains *D*-phenomena independently of other theories, and *T* does not explain anything non-*D* without being embedded in a larger theory. *T* is explanatorily invariant over *D*. (##7)

Deploying this notion, I think that grammars primitively explain linguistic expressions not mental states. In contrast, Collins thinks: "A linguistic theory *primitively explains* the speaker-hearer's understanding, not properties of the inscriptions themselves" (##16); "A grammar is supposed to explain the character of our capacity to produce and consume material *as* linguistic" (##13). But how could it do that?

Here are some typical grammatical rules (principles):

An anaphor must be bound by another expression in its governing category.
A pronoun must not be bound by another expression in its governing category.
Accusative case is assigned by a governing verb or preposition.
A verb which fails to assign accusative case fails to theta-mark an external argument.

These are about *expressions*. They are not about mental states: they *do not mention* understanding or mental capacities (31). Building on this, in correspondence with Rey and with reference to Quine (1961), I offered the follow deductive argument for my view of grammars, step (4) above (in Collins' numbering):

(a) Any theory is a theory of x's iff it quantifies over x's and if the singular terms in applications of the theory refer to x's.

(b) A grammar quantifies over nouns, verbs, pronouns, prepositions, anaphors, and the like, and the singular terms in applications of a grammar refer to such items.[7]
(c) Nouns, verbs, pronouns, prepositions, anaphors, and the like are linguistic expressions/symbols (which are entities produced by minds but external to minds).
(d) So, a grammar is a theory of linguistic expressions/symbols.[8]

Now one might well think that grammatical rules *directly* explain the natures of expressions but also *indirectly* explain something mental, linguistic competence. As Collins has remarked earlier, "the theory could tell us about speaker-hearers by way of being true of something else, where what a theory is true of should attract our proper ontological commitment" (##3–4). Why does Collins not think that what *could* thus be the case is *actually* the case?

In a response to Collins I attempted to show how it *is* actually the case: "It is *because* the grammar gives a good explanation of the symbols that speakers produce that it can contribute to the explanation of the cognitive phenomena" (2008a: 215). I illustrated this with an example I took from Collins (slightly modified here to fit Collins' present discussion). The cognitive phenomenon to be explained is S's interpretation of the reflexive in *Fred's brother loves himself* as referentially dependent on the whole DP rather than on *Fred* or *brother* alone. Collins nicely reproduces my answer as follows:

(EE)
(i) S is competent in English and hence *respects* its structure rules.
(ii) *Fred's brother loves himself* is an English sentence in which *himself* is c-commanded by the whole DP but not by either of its constituents.
(iii) It is a rule of English that, in these circumstances, the reflexive must be bound by the whole DP.
(iv) Therefore, S, because he respects the rules of English, gives a joint construal to *himself* and the whole DP. (##20)

Notice that the only psychological premise in this explanation is (i), what I call "the minimal claim on the issue of the psychological reality of language" (25). And notice that the explanation is not deep. It does not tell us much about S's mind because (i) is *so* minimal: the explanation just tells us that, *whatever it may be that constitutes her competence in English* causes S to interpret a reflexive according to the rules of English. That is hardly news. I continued on:

> This cognitive explanation depends on (ii) and (iii) providing a good linguistic explanation of the nature of that English sentence. *In general, English speakers construe English*

[7] For a discussion of some examples in Liliane Haegeman's textbook (1994), see my 2008b: 250–251.

[8] Rey writes (##8) as if *this* is my argument for the linguistic conception: it is only the final move, step (4). Later, he curiously remarks that I don't "present anything like a serious theory of external 'Linguistic Reality' remotely comparable in richness and power to an internalist generative grammar" (##12). But, of course, a central point of the linguistic conception is that *those very grammars are*, precisely, rich and powerful theories of external linguistic reality.

> *expressions as if they had certain properties because, as the grammar explains, the expressions really have those properties.* (2008a: 215).

Finally, the grammar directly explains the linguistic phenomenon of reflexives "independently of other theories" – see (ii) and (iii) – but it does not so explain the psychological phenomenon of S's interpretation of reflexives – see (i). So, the grammar "primitively explains" the linguistic not the cognitive phenomenon.

19.1.1.4 The Paraphrase Response

Return to Colllins. Why does he think that "the grammar tells us nothing about the strings, but lots of things about how we interpret them" (##16)? *Prima facie*, this is all wrong, as argument (a) to (d) above shows: as I put it, "the Chomskian research program is revealing a lot about language but, contrary to advertisements, rather little about the place of language in the mind" (16). But is this view of grammars a bit flatfooted? As Quine points out, it is quite respectable for a scientist to paraphrase by withdrawing claim *P* in favor of claim *Q* because *Q* lacks the ontological commitment of *P* and yet will serve his purposes well enough. The scientist thus "frees himself from ontological commitments of his discourse" (1961: 103). The commitments of *P* arose from "an avoidable manner of speaking" (13). So, as Rey says,

> we shouldn't go merely by what is said in elementary textbooks, but, rather, what sort of entities are performing the *genuine explanatory work* the theory is being invoked to perform (##10)

In light of this paraphrase response, enthusiasts for the psychological conception need to do three things. First, *they need to acknowledge* that grammars should *not* be taken literally but rather should be taken as standing in for a set of paraphrases that do not talk about expressions (like reflexives and DPs). Second, *they should tell us what the paraphrases are* or, at least, tell us how they are to be generated from what grammars actually say. Third, *they should give examples of how this paraphrased grammar directly explains cognitive phenomena*. So far as I know, nobody, from Chomsky down, including the critics of my linguistic conception,[9] has met the last two requirements explicitly in print. So we should be grateful to Collins for meeting them in his present paper.[10] Given the centrality of the psychological conception to the promotion of Chomskian linguistics, it is striking that there has been

[9] See note 5.

[10] Later, however, Collins seems to back away from the need to paraphrase. He raises the question: "How would linguistics appear if linguists were not interested in external concreta, but in the abstract mental structures speaker-hearers employ in their interpretation and production of such concreta, *inter alia*?" He responds "Well, it would appear just as it does, surely" (##26). But this is not so: grammars quantify over linguistic expressions and these are not "abstract mental structures".

so little sensitivity to these requirements in presentations of grammars. Sensitivity to the ontology of one's theory is a mark of good science.

So how does Collins meet the requirement of the paraphrase response? He rejects my explanation, (EE), and gives an ingenious "internalist" explanation "without appeal to rules of English or external properties". This explanation is very revealing:

> (IE)
>
> (i) If S is competent in English, S's interpretation of the marks *himself* is constrained to be jointly interpreted with the interpretation of other marks occurring in the inscription i.e., the interpretation is 'reflexive'.
>
> (ii) The constraint is such that the interpretation of *himself* is dependent on the interpretation of a mark categorised by S as a c-commander of the first interpretation.
>
> (iii) To determine a c-commander, S must project the lexical interpretations into a hierarchical structure determined by the interpretations mapped onto the given marks.
>
> (iv) Based on the projection, the interpretations of *Fred* or *brother* do not c-command the reflexive interpretation; only the interpretation of *Fred's brother* does.
>
> (v) Therefore, the reflexive is jointly construed with the DP interpretation. (##21)

First, some clarification. (IE)'s talk, in its steps (i) to (iv), of S's interpretation of marks "does the work" of (EE)'s talk, in its step (ii), of the properties of the external marks themselves. Those steps (i) to (iv) establish that if S is competent with English then *S interprets* the mark *himself* as a reflexive that is c-commanded by *S's interpretation* of the whole DP mark *Fred's brother* but not by *S's interpretation* of either of its parts. But what then "does the work" of (EE)'s (iii)? How do we get from (IE)'s first four steps to its conclusion (v)? There is an implicit premise along the following lines:

> (Ri) When speakers competent with English interpret a mark as a reflexive and c-commanded by a DP mark but not by either of the DP's constituents, they jointly construe the reflexive with the DP interpretation.

We should start thinking about this proposal by reminding ourselves that Collins needs to show that *the grammar* "primitively explains the speaker-hearer's understanding". So, *where is the grammar playing a role in (IE)*? Manifestly, the grammar of reflexives *as presented in textbooks* is not playing a role. For, that bit of the grammar, reflected in (EE)'s (iii), is along the following lines:

> (Re) In an English sentence in which a reflexive is c-commanded by a DP but not by either of the DP's constituents, the reflexive is bound by the whole DP.

For Collins to be right that the *grammar* primitively explains S's interpretation, the textbook grammar containing (Re) has to be replaced by one containing (Ri): *a grammar's apparent commitment to external linguistic expressions has to be paraphrased away into a commitment to acts of interpretation*. Similarly, (EE)'s grammatical claim (ii) about the c-commanding of expressions has to be paraphrased away as a grammatical claim about the interpreting of expressions as c-commanding, based on (IE)'s (i) to (iv). Collins, in effect, acknowledged the need for this sort of paraphrase earlier:

> When a linguistics text tells us that a sentence is ambiguous, say, the claim is not that the exemplified string has some peculiar hidden structure or some high-level functional prop-

> erty or any other property as an external entity. The claim is simply that competent speaker-hearers robustly and reliably associate 2+ specific interpretations with tokens of such a string type. (##18–19)

Collins' explanation (IE) is indeed one "without appeal to rules of English"; instead, it appeals to rules governing interpretations by English speakers.

What is Rey's story? His examples of cognitive phenomena that need to be explained are of speakers finding certain constructions – "WhyNot?s", as he neatly names them – "oddly unacceptable" (##10). He claims that it is "virtually impossible to imagine a non-psychological explanation of these phenomena" (##11). And so it is, because the phenomena *are* psychological! But the issue is whether *grammars* directly explain them. I would argue, along the lines of (EE), that grammars as they stand do not directly explain them, although their accounts of the syntactic structure of linguistic entities contribute to the explanation.[11] Rey needs an alternative explanation provided by a suitably paraphrased grammar. His suggestion for paraphrase, illustrated with the principle for negative polarity items, is to replace principles about expressions with principles about (mental) *representations* of expressions (##13). This has the great advantage for Rey of removing a grammar's ontological commitment to linguistic expression and hence to linguistic realism. We should be grateful to Rey, as to Collins, for proposing a paraphrase. But, we should note first, that Rey's paraphrase is *very* different from that of Collins in terms of the interpretation of marks, along the lines of (Ri). And, we should note second that it is unclear how Rey's paraphrase will yield the desired psychological explanations. How, for example, will it explain speakers' finding WhyNot?s "oddly unacceptable"? Rey follows up his paraphrases with another suggestion that seems more explanatorily promising: he recommends "simply to *pretend* there are [linguistic items] and tree-structures in which they appear, and then treat any rules, principles, constraints, or parameters as true under the *pretense*" (##13). If this were a story of what goes on when speakers interpret marks as linguistic entities then it would be similar to Collins's story.

19.1.1.5 Criticism of the Paraphrase Response

1. Collins has offered a story of how the grammar directly explains cognitive phenomena. And if the grammar were rewritten along the lines of (Ri), it seems that it would indeed explain psychological phenomena. But note that such explanations can be derived from the *un*revised grammar together with the uncontroversial claim that competent speakers "respect" the principles and rules of the grammar; see (EE). *The rewriting adds no explanatory power*. And the rewriting

[11] To avoid misunderstanding, perhaps I should emphasize that I grant the Chomskian view that the language has some of its syntactic properties, including perhaps the WhyNot?s, as a result not of convention but of innate constraints on the sorts of language that humans can learn "naturally" (2006a: 244–272).

is pointless if the grammar is indeed a more or less true account of linguistic reality, as the linguistic conception claims.

2. Collins – and perhaps Rey – is proposing, in effect, that a grammar's claim that external marks have linguistic properties should be paraphrased as the claim that speakers interpret them *as if* they had those properties. This has a sadly nostalgic air to it. The idea that widely accepted claims about the nature of the external world should be rewritten as claims that it is *as if* that world has that nature – that it *seems to us* that it does – has a long grim history that includes the disasters of metaphysical idealism and scientific instrumentalism.[12] In my view (1984, 1991), the idea has little to be said for it.
3. If Collins' and Rey's approach to linguistic entities were good it would generalize to social entities in general. So if their argument really showed that linguistic entities were "explanatorily otiose" then analogous arguments would show that, for example, votes and coins were "explanatorily otiose": all we need for explanatory purposes is the fact that people *treat* certain pieces of paper *as if* they were votes and certain pieces of metal *as if* they were coins. But this is not so. Those pieces of paper play their causal roles in our lives *in virtue of* having the property of being votes; that's what explains their place in the causal nexus. Similarly, those pieces of metal. And similarly, the marks we treat as linguistic entities. Nothing in Collins' or Rey's discussion, it seems to me, gainsays this causal role of linguistic entities. In particular, what explains the fact that *everyone* in the English-speaking community would interpret the mark *himself* in *Fred's brother loves himself* as a reflexive c-commanded by the mark *Fred's brother*? Why this remarkable coincidence of interpretation? The core of the answer is simple: because in English *himself is* a reflexive c-commanded by that DP[13]; and the competence of any English speaker enables her to detect this linguistic fact; that's part of *what it is* to be competent. There are similar simple answers to analogous questions about other social entities like votes and coins. *We suppose that certain parts of the external physical world have certain natures because having those natures explains their causal roles.* There is no good reason to resist these answers. Social "as-if-ism" is a mistake.

Collins makes an Occamist claim: "The crucial issue here is parsimony" (##21). I agree. Linguists, like other scientists, should only posit what they need for explanation. We need linguistic entities for that purpose just as we need votes and coins. There are, external to the mind, entities playing causal roles in virtue of their linguistic properties. Grammars are approximately true theories of some of those properties, the syntactic ones.

[12] Rey firmly rejects this sort of antirealism (##13 n. 28) but one wonders why the external world in general is spared the treatment he gives external languages in particular.

[13] Collins begins a paragraph with the charge that my "position is incoherent as it stands". How so? He sums up: "Suffice it to say, nowhere does Devitt attempt any analysis of the properties he mentions to render them as high-level functional properties of external marks" (##26–27). I take over my view of these properties from grammars. What more "analysis" do I need?

19.1.2 The Psychological Reality of Language (Camp)

After my argument in Chap. 2 for the first major conclusion, the linguistic conception of grammars, the rest of *Ignorance* is largely devoted to this question: In what respect, if any, are the structure rules of a language "psychologically real" in its competent speaker? Given the linguistic conception we know that the speaker's competence must "respect" those rules: there must be something-we-know-not-what within the speaker apt to produce outputs governed by those structure rules; analogously, to process inputs. But what is that something? And how much of it is innate? My struggle with these questions yields the remaining five of my seven major conclusions and all seven of my tentative proposals. So my answers are complicated! This reflects how very difficult it is to tell what they should be: "A central theme of the book is that we have nowhere near enough evidence to be confident about many psychological matters relevant to language; we are simply too ignorant" (vi). Hence the tentative nature of many of my proposals. They are put forward not to be believed but as promising hypotheses to be further investigated.

In "Priorities and Diversities in Language and Thought", Elizabeth Camp agrees with me about the linguistic conception: "I share Devitt's conception of language as a public, social construction" (##3). But she disagrees with me over thoughts and the psychological reality of language.

My views on these matters begin with "intentional realism" – there really are thoughts (125–127) – and "LET" – "language expresses thought" (127–128). This leads to the view that:

> linguistic competence is an ability to match sounds and thoughts for meaning. If this is right then it is immediately apparent that any theory of linguistic competence, and of the processes of language comprehension and production, should be heavily influenced by our view of the nature of thoughts. (129)

This leads to:

> *Fourth major conclusion*: The psychological reality of language should be investigated from a perspective on thought. (129)

I go on to argue for four priority claims (125–141) which I sum up:

> *Fifth major conclusion*: Thought has a certain priority to language ontologically, explanatorily, temporally, and in theoretical interest. (141)

On the basis of these conclusions I go on to propose some hypotheses about thoughts and the psychological reality of language.

Camp disagrees with both the ontological and explanatory priority claims. I shall consider only the explanatory one.[14] She claims that "neither thought nor language can be assigned clear explanatory priority over the other" (##3). My argument that

[14] I considered an earlier version of her disagreement with my ontological claim in my responses at the 2007 Pacific APA; also, an earlier version of her disagreement over the language faculty.

thought is prior comes from Paul Grice (1989), building on LET. The crux of the argument is as follows:

> An utterance has a "speaker meaning" reflecting the meaning of the thought that the speaker is expressing, reflecting its "message"…. [S]peaker meaning is explanatorily prior to conventional meaning: the regular use in a community of a certain form with a certain speaker meaning leads, somehow or other, to that form having that meaning literally or conventionally in the language of that community. So the story is: thought meanings explain speaker meanings; and speaker meanings explain conventional meanings. In this way thought is explanatorily prior to language. (132).

So where does Camp disagree?

Camp has a nice discussion in Sect. 3.2 in support of the view that "language does much more than express thought" (##3). She sees this as a rejection of LET. It is not at all clear to me that it is at odds with LET as I understand it or, more importantly, with the development of LET in the Gricean argument above. In any case, what clearly is central to Camp's disagreement is her rejection of "the language of thought hypothesis" (LOTH), the view that thoughts involve language-like mental representations. In Sect. 3.1, Camp presents some very interesting and helpful considerations in favor of the view that "humans employ a range of formats for thought" (##3). But the way she brings this rejection of LOTH to bear on my discussion reflects serious misunderstandings. Indeed, her attributions of views to me are frequently quite wide of the mark, overlooking many of the complications of my treatment. I have space here to discuss just two.

I begin with the following:

> The first step in Devitt's broadside for the priority of thought is establishing that thought itself has a sentential structure…. [That thesis] plays a central role in his overall argument for the priority thesis. (##4).

But, in fact, LOTH *plays no role at all* in my case for the explanatory priority of thought; see the Gricean argument. Indeed, LOTH does not enter my discussion until the following chapter. Furthermore, Camp overstates my commitment to LOTH. I do indeed present an argument for LOTH (145–147, 158), influenced by Jerry Fodor (1975), and I do favor it over alternatives, but I never claim to have *established* it. I regard LOTH as "far from indubitable" (11), a "fairly huge" step (148), "a bold conjecture" (259), "speculative" (260), an example of a "promising" theory (192) to be further investigated. So *none* of my main "conclusions" depend on it. True, some of my seven other "proposals" do depend on it, and that is sufficient for my labeling them "tentative".

The second misunderstanding I shall discuss concerns, in effect, one of those proposals:

> *First tentative proposal*: A language is largely psychologically real in a speaker in that its rules are similar to the structure rules of her thought. (157).

I summed up the argument for this (142–162) as follows:

> First, we adopted the controversial LOTH. Second, we noted, what seems scarcely deniable, that public language sentences can have syntactic properties that are not explicit.

> Third, we pointed to the role of mental sentences in explaining the implicit and explicit syntactic properties of the public sentences that express them. (157–158).

I noted that “the importance of LOTH to this argument is enormous” (158).

Camp is way off track about this discussion. She talks of my “ultimate conclusion that language exhibits the particular features it does because thought possesses those feature” (##4). But this hypothesis of in virtue of what language has its syntax is not an “ultimate conclusion” at all: it is part of an argument for a tentative proposal; see above. Later she alleges that “the next big move in Devitt’s argument for the priority of thought is the claim that language takes the form it does because it expresses thought” (##9). But that claim about language has nothing to do with my priority argument, which is in the previous chapter and is part of the background for this discussion of the first tentative proposal. So, Camp has the order of argument backward.

We are investigating the psychological reality of language. My first tentative proposal is my best guess *if the controversial LOTH is true*. But I appreciate that it may well not be and so investigate the consequences of that (195–243). I argue for:

> *Fifth tentative proposal*: If LOTH is false, then the rules of a language are not, in a robust way, psychologically real in a speaker. (243).

So, if Camp is right in rejecting LOTH, I would conclude something strongly at odds with the standard Chomskian view of the psychological reality of language. I wonder whether Camp embraces that anti-Chomskian conclusion.

I also argue (256–260) for:

> *Sixth tentative proposal*: Humans are predisposed to learn languages that conform to the rules specified by UG because those rules are, largely if not entirely, innate structure rules of thought. (257)

This is the innateness analogue of my first tentative proposal. And Camp’s critical discussion of it has, I would argue, misunderstandings analogous to her discussion of the first. But if she is right about LOTH, then this proposal goes down with the first. Appreciating that it might, I argue (256–270) for an analogue of the fifth:

> *Seventh tentative proposal*: If LOTH is false, then the rules specified by UG are not, in a robust way, innate in a speaker. (260).

So we would again have a conclusion strongly at odds with a standard Chomskian view. I wonder whether Camp embraces that conclusion.

In sum, put Camp’s interesting argument against LOTH together with my arguments for the fifth and seventh tentative proposals and we have a case against the received Chomskian view that what grammars and UG describe is psychologically real in speakers.

19.2 Theory of Reference

19.2.1 Reference Borrowing (Raatikainen, Sterelny, Horwich, Recanati)

Saul Kripke made two enormous contributions to the theory of reference, one negative, one positive. The negative one was what I call "the ignorance and error argument" against description theories of reference. This showed that many competent users of many terms, most strikingly proper names, use those terms successfully to refer to entities about which those users are largely ignorant or wrong. So, the reference of those names could not be determined by the descriptions that those competent users associate with the terms as standard description theories require.[15] The positive one was that people can "borrow" the reference of many terms in a communication situation, again most strikingly proper names, in an epistemically undemanding way that does not require any capacity to identify the referent. This radical idea is the crux of Kripke's "better picture" of reference (1980: 94).

Kripke's reference-borrowing idea for names has frequently been misunderstood. Panu Raatikainen points out, in his informed and judicious "Theories of Reference: What Was the Question?":

> Both the initial borrowing and the later use are intentional actions, but [the] … subsequent use need not involve any intention to defer to the earlier borrowing; it need not involve any "backward-looking" intention. (##8)

John Searle (1983: 234) and others have wrongly taken Kripke's reference borrowing to require such a backward-looking intention and have, perhaps for this reason, renamed it "deference", as I have noted (2006e: 101–102; 2011a: 202–204). What matters for the subsequent use of a name that has been acquired by reference borrowing

> is that the use be caused by an ability with that name that is, *as a matter of fact*, grounded in the bearer via that borrowing: *the efficacious ability must have the right sort of causal history*… [T]he speaker need not know who the lender was or even that she *has* borrowed the name. There is no need for her to have any semantic thoughts about the name at all. Use of language does not require any thoughts about language. (2015b: 116)

What does the initial borrowing require? I have, as Kim Sterelny nicely puts it in "Michael Devitt, Cultural Evolution and the Division of Linguistic Labour", a "minimalist position" on this (##2). I do not follow Kripke in talking of the borrower intending to use the name with the same reference as the lender (1980: 96). That strikes me as too intellectualized. I recently put it like this in "Should Proper Names Still Seem So Problematic?" (2015b), the most accessible and up-to-date version of my theory:

[15] As Raatikainen point out (##2), this argument counts against theories that require identification by description *or* ostension.

> Rather we require that the borrower process the input supplied by the situation in whatever way is appropriate for gaining, or reinforcing, an ability to use the name to designate its referent. The borrower must intentionally set in motion this particular sort of mental processing even though largely unaware of its nature and perhaps not conscious of doing so. Similarly, a person walking or talking must intentionally set in motion the sort of mental processing appropriate to that activity. (2015b: 117–118)[16]

In “Languages and Idiolects”, Paul Horwich criticizes an earlier presentation of this view of reference borrowing (Devitt 2011a: 202):

> this amounts to saying that only when certain unspecified conditions are satisfied will a speaker’s term refer to what the person she heard it from was referring to. But this is so uninformative that we as well might impose no extra condition at all. (##7)

This criticism would be appropriate only if we should expect philosophers to specify these conditions. But we should not[17]: “Reference borrowing is a species of lexical acquisition or understanding and so we must look to psycholinguistics to throw more light on it” (2015b: 118). And given the current state of our knowledge of language processing, we should not expect psycholinguists to tell us much soon. Sterelny hypothesizes “that rich and cognitively complex subdoxastic capacities are needed” for reference borrowing (##12). That seems likely to me.

In “Multiple Grounding”, François Recanati (##1) quotes my saying of names that “reference borrowing is of the essence of their role’ (Devitt 1981a: 45). This essentialist claim is a bit misleading. Searle (1983: 241) has objected to Kripke that it is not essential that people borrow their reference of a name; each person might manage reference on her own. I responded: “No causal theorist has ever denied this. The reference of any name *can* be borrowed, even if those of a few are not in fact borrowed” (1990: 101–102 n. 9). That is how to understand my essentialist claim.

Still, a person’s reference with a name *is* typically borrowed, at least partially. I pointed out that this borrowing is like pronominal anaphora (1981a: 45). Recanati puts it nicely:

> just as the pronoun ‘he’ inherits the reference of its singular antecedent … in the dialogue “Have you read Aristotle? Yes, *he* is a great philosopher”, the name ‘Aristotle’ in [this sentence] inherits its reference from past uses to which that use is causally related. (##2)

Recanati thinks that with both the anaphoric link for pronouns and the “quasi-anaphoric” link for names, the

> link forces coreference between the singular term (name or pronoun) and the singular terms it is linked to in the chain or network. Coreference, in these cases, is more or less mandatory. It is *de jure*, not *de facto*. (##7)

[16] A similar theory of reference borrowing is also appropriate for referential descriptions (Devitt 1974: 191–192; 1981a: 38–39). David Kaplan has recently also urged such a view (2012: 142, 147).

[17] Perhaps I should help myself to Horwich’s concluding remarks about his own view:

Moreover, even if the account presented here would be somewhat improved by the elimination of some of its present imprecision … it’s quite possible that, even as it stands, the view is cogent, correct, and an illuminating step in the right direction. (##13)

The cautious "more or less" is important here. For, only coreference demanded by the syntax is *really* mandatory. Thus, consider 'John thinks that Bruce loves himself'. In any utterance of this sentence, 'himself' *must* corefer with 'Bruce'. In contrast, it is not the case that in any utterance of 'Yes, *he* is a great philosopher' in response to 'Have you read Aristotle?' 'he' *must* corefer with 'Aristotle''. It *probably* will, of course, because the speaker will probably be expressing a thought that it is grounded via an anaphoric link to the earlier 'Aristotle'. Still, it might not corefer with 'Aristotle' because the speaker might be expressing a thought that is grounded via an anaphoric link to another singular term earlier in the discourse; or she might be expressing a "demonstrative" thought using 'he' deictically. Nor is it the case that any utterance of 'Aristotle' must corefer with any particular earlier token of 'Aristotle', say, one referring to the famous philosopher. Whether the utterance does depends on whether the speaker is expressing a thought that is grounded via a reference borrowing link to that earlier token; the thought expressed might be grounded in Onassis via a different link.

Still, it is true that *if* the speaker's token is expressing a thought grounded (solely) via an anaphoric or quasi-anaphoric link to another token *then* the two tokens must corefer. And it is also true that speakers have a social obligation not to mislead. So their utterances should exploit the links that an audience will naturally take them to be exploiting. Relatedly, Recanati notes that "in general, use of *the same word* by the interlocutors triggers a presupposition of coreference" (##7). Nonetheless, that presupposition may be false: the presupposition does not *make it the case* that there is coreference; a speaker may be carelessly, or even deliberately, misleading. As I have emphasized elsewhere (2013c), what constitutes reference is one thing, what hearer's reasonably take the reference to be is another.

Just how extensive is reference borrowing? Kripke took it to be a feature not only of names but of "natural" kind terms. And Hilary Putnam went even further with his "division of linguistic labor" (1973, 1975). Raatikainen believes that "any sort of term can be borrowed" (##11), even 'bachelor' (##10). I wonder. Consider "artifactual" kind terms. It does seem plausible that the reference of terms like 'sloop' and 'dagger' can be borrowed. But if so, are they, like proper names, covered by a "pure-causal" theory of borrowing? Sterelny and I pointed out in our textbook that maybe not (1999: 93–101): although you can gain 'sloop' without associating it with 'boat having a single mast with a mainsail and jib', maybe you cannot gain it without associating it with 'boat'. And what about the more basic "artifactual" terms 'boat' and 'weapon'? Is it really plausible that they can be borrowed? The first problem with these questions is that we do not have strong intuitions. The second problem is that here, more than anywhere in the theory of reference, we need the support of more than intuitions. We need evidence from usage (2011b, 2012a, b, c, 2015c). We shall consider how to get this in Sect. 19.4.4.

19.2.2 Grounding (Raatikainen, Recanati)

As a result of reference borrowing we can all succeed in designating Aristotle with his name in virtue of a chain of reference borrowings that takes us back to the original users who fixed the name's reference in Aristotle. But how did the original users do that? I have urged that the reference of a *paradigm* name like 'Aristotle' is fixed in the object, directly or indirectly, by the causal link between a person and that object when it is the focus of a person's perception. This is what I call a "grounding" (1974: 185–186, 198–200; 1981a: 26–29, 56–64, 133–136; 2015b: 113–115). (We shall consider the reference fixing of *non*-paradigm "descriptive" names like 'Jack the Ripper' in Sect. 19.3.2.) The grounding may be direct, by formal, informal, or implicit dubbing; or it may be indirect, via other terms that are grounded in the referent. I wanted my theory, as Raatikainen aptly remarks, to be "thoroughly causal, also at the stage of introduction of names" (##6).

According to the theory of grounding, a name is grounded in an object by a certain causal-perceptual link. This sort of situation will typically arise many times in the history of an object leading to my claim that names are typically *multiply grounded* in their bearers (1974: 198). I used this idea, together with Hartry Field's (1973) idea of partial reference, to explain cases of reference confusion (1974: 200–203). Thus, applying these ideas to Kripke's famous raking-the-leaves example (1979a: 14) yields the conclusion that 'Jones' has a semantic-referent, Jones, but no determinate speaker-referent; both Jones and Smith are *partial* speaker-referents (1981b: 512–516; 2015b: 118–121). Later, I applied the ideas to cases of reference change like Gareth Evans' famous example of 'Madagascar' (1981a: 138–152; 2015b: 121–124). In brief, the reference of a name changes from *x* to *y* when the pattern of its groundings changes from being in *x* to being in *y*. Nonetheless, the mistaken idea that cases of reference change are "decisive against the Causal Theory of Names" (Evans 1973: 195) persists (Searle 1983; Sullivan 2010; Dickie 2011).[18]

Recanati suggests that the "theory of multiple grounding can … be construed as a response to the challenge raised by Evans for Kripke's theory (##3). It was certainly used as a response but it was not *introduced* for that purpose; it is a consequence of the causal-perceptual theory of groundings as I emphasized in "Still Problematic":

> Groundings fix designation. From the causal-perceptual account of groundings we get the likelihood of multiple groundings. From multiple groundings we get the possibility of confusion through misidentification. From confusion we get the possibility of designation change through change in the pattern of groundings. (2015b: 123–124)

Recanati demonstrates convincingly that the ideas of multiple grounding and partial reference can be applied to pronouns (##5–9). And he rightly thinks that his "mental

[18] Multiple grounding is also vital in explaining reference change in "natural kind terms" (1981a: 190–195). So, as Raatikainen points out (##13), it can deal with "most of Unger's much-cited alleged 'counterexamples'" to the causal theory (Unger 1983).

file framework … is, by and large, compatible with Devitt's causal account of reference" (##14); indeed, I came to explicitly adopt something like it (1989a, 1996).

19.2.3 Kripkean or Donnellanian? (Bianchi)

In an earlier paper with Alessandro Bonanini (2014), Andrea Bianchi argues that the pictures of the reference of proper names given by Donnellan and Kripke are different: where Donnellan thinks that reference is determined by a cognitive state of the speaker not the social mechanism of reference borrowing, Kripke thinks the opposite.[19] In "Reference and Causal Chains", Bianchi alleges initially that my position is unclear in that I confuse Donnellan's and Kripke's pictures (##12). He goes on to argue that, in the end,

> Devitt's causal theory shouldn't be seen as a development of Kripke's chain of communication picture. It should, rather, be considered as a development of the alternative causal picture offered by Donnellan, Donnellan's historical explanation theory. (##17)

I think that Bianchi is wrong about all three of us. Donnellan's and Kripke's pictures of reference determination, charitably understood, *both* include *both* a cognitive state and reference borrowing. My theory certainly does. So it can be considered a development of both pictures.

We should not lose sight of the fact that the most important question about any theory is whether it is *true*, whatever its origins. Beyond that, credit or blame for a theory should of course be placed where it is due. Now, as I have always made clear (1974: 184 n. 2; 1981a: x–xi), my theory of names was influenced by Kripke's picture not Donnellan's[20]; indeed, my first presentation of the theory, in my dissertation (1972), was completed before I read Donnellan (1970) on names (which was mentioned in a last-minute footnote). I took Kripke's picture of reference borrowing from his 1967 Harvard lectures and developed it in a naturalistic setting (of which Kripke would not approve, of course); see Sect. 19.2.1. I added to this a causal theory of reference fixing ("grounding"), summarized in Sect. 19.2.2, for which Kripke should not, of course, be blamed.[21] However, my main concern here is not to

[19] Antonio Capuano (2018) and Julie Wulfemeyer (2017) agree.

[20] Donnellan did, however, have a big influence on my view of descriptions. I urged a causal theory of referential descriptions (1972, 1974: 190–6; 1981b) as well as of names and demonstratives, as Bianchi notes (##5). This theory rests on a referential/attributive distinction I drew (1981a: x–xiii) under the influence of Donnellan and C.B. Martin.

[21] Bianchi and Bonanini take me, in my text with Sterelny (1999: 66–67), to ascribe "a causal theory of reference borrowing (*as well as one of reference fixing*)" to both Kripke and Donnellan (2014: 195 n. 37; emphasis added). Now what we actually ascribe to them is "*the basic idea* of causal, or historical, theories of reference" (1999: 66; emphasis added). We then go straight on to present "our theory" which includes a causal theory of *both* fixing and borrowing. So it is natural to take our ascription to Kripke and Donnellan to be of both. That was misleading: the basic idea we intended to ascribe was the theory of borrowing. Indeed, in earlier works, I ascribe a causal

argue about what my theory of names developed from but to clarify and defend the theory.

A startling dialectic has emerged, largely out of UCLA, that is a helpful background to understanding Bianchi's claims. The dialectic is on fundamental issues about language. On one side there is Bianchi himself, in the wonderfully provocative "Repetition and Reference" (2015). He rejects the near-universal Gricean view that mental content is explanatorily prior to linguistic meaning:

> in my opinion things should be the other way round: the intentional properties of (post-perceptual) mental states are to be explained in terms of the semantic properties of linguistic expressions. (2015: 96)

This view is reflected in Bianchi's austere account of reference borrowing. He eschews Kripke's talk of intentions (Sect. 19.2.1 above), claiming "that this problem can be dealt with by appealing to the notion of *repetition*" (2015: 100), a notion he takes from David Kaplan (1990). On the other side of the dialectic, there are some philosophers who Bianchi calls "neo-Donnellanians", responsible for "a Donnellan *Renaissance* in the theory of reference" (##11). The seminal work in this Renaissance is a volume of papers, *Having in Mind*, edited by Joseph Almog and Paolo Leonardi (2012). The neo-Donnellanians reject reference borrowing and hold that a person's use of a name refers to whatever she has in mind in using it. I am dismayed by Bianchi's conclusion that I belong with them.

For reasons implicit in my "What Makes a Property 'Semantic'?" (2013d), I think both sides of this dialectic are deeply wrong. (Is there something in the UCLA water?) Against Bianchi, a commitment to the Gricean priority runs right through my work; see Sect. 19.1.2 above, for example. The contents of thoughts *create*, *sustain*, and *change* the meanings of linguistic expressions. So far as reference borrowing is concerned, we have already noted that I have what Sterelny calls a "minimalist position" (Sect. 19.2.1). Still, my position is a lot more robust than Bianchi's, which is far too weak to do the job in my view. Reference borrowing does not require intentions[22] but it does require mental processes, probably complicated subdoxastic ones, to be discovered by psycholinguists. I shall say no more about Bianchi's position. My discussion of Donnellan will give some indication of why I disagree with the neo-Donnellanians.[23] For, Bianchi's view that Donnellan's picture differs from Kripke's is based on the surprising interpretation that Donnellan subscribes to the above neo-Donnellanian view.

In Bianchi's present paper, he sums up his earlier argument with Bonanini (2014) as follows:

theory of reference borrowing to Kripke as "the central idea" (1972: 55 n. 4; 1974: 184), "the vital feature" (1972: 73), of the causal theory. I saw the causal-perceptual theory of grounding as my contribution (1972: 55 n. 4; 1981a: xi). My take on Kripke's view of reference fixing is set out in "Still Problematic" (2015b: 113).

[22] Indeed, in "Three Mistakes About Semantic Intentions" (Forthcoming-a) I argue that there is no place at all for talk of intentions in semantics.

[23] See also the criticisms in Martí 2015.

> while reference borrowing is obviously fundamental for Kripke, there is no place for this alleged phenomenon in Donnellan's account…. [W]hen we use a proper name we simply do not borrow reference. On the contrary, we always *fix* it *anew*. (##9)

Bianchi and Bonanini's case for this is ingenious, and based on a careful study of Donnellan. Nonetheless, I think that it is probably wrong. Consider the name 'Aristotle'. This name is frequently used by philosophers to refer to a certain famous ancient Greek philosopher, as everyone would agree. Far from fixing its reference anew, these philosophers do something that has been done countless times for centuries. They use the name to refer to that ancient philosopher by *participating in a convention* of using the name to refer to him; their usage is governed by a *linguistic rule* that links the name to the philosopher; as linguists say, 'Aristotle', with that meaning, is an *item in their lexicon*. Rules/items like this, established by conventions (apart from some innate syntax; Sect. 19.1.2), constitute our language. And to deny that our uses of expressions are (largely) governed by such rules is, in effect, to deny that we *have* a language. Bianchi draws attention to this consequence of his interpretation:

> extreme consequences … can be drawn from Donnellan's account, for example that, at least from a semantic point of view, there are no *languages*. (##17–18)

I first thought that no sensible philosopher like Donnellan could embrace anything close to this. But then I started to read the neo-Donnellanians.[24]

Consider Jessica Pepp, for example. In a very rich paper (2018), she rejects what I have just urged, which she calls "the *Conventional Stance*", the thesis that "the reference of proper names as used on particular occasions is determined by linguistic convention" (##4).[25] After arguing effectively against Bianchi's repetition account, she asks: "If participating in a convention for using a proper name is not some form of copying or repeating previous uses, what is it?" (##9). She is rightly critical of intention-based answers and finds no satisfactory answer. But she does not consider a less intellectualized approach focused on largely non-central causal processes. On my account,

> for there to be a convention in a community of using an expression with a certain speaker meaning is for members of the community to be disposed to use that expression with that meaning *because other members are disposed to do so*: there is a certain sort of causal dependency of the disposition of each member of the community on the dispositions of others. And for a person to participate in this convention is for her to use that expression *because* she has that dependent disposition. (2015b: 119)

[24] I am indebted to Bianchi for directing my attention to these works and hence awakening me from my "dogmatic slumbers".

[25] But, a word of caution about the tricky word 'determined'. In this thesis it should have its causal sense not its epistemic or constitutive sense: a convention causes there to be a linguistic rule. The linguistic rule, whether innate, idiosyncratic, or caused by convention, is what constitutes the reference of the name (2013d: 96–100; Forthcoming-b). As a matter of fact, the rule for 'Aristotle' and, say, 'dog' were caused by convention.

A language is constituted by rules for its expressions. People are speakers of the language in virtue of being disposed to use its expressions in accordance with those rules, dispositions that are mostly established by convention (see Sect. 19.3.5 for more). It is hard to say precisely what is involved in a person exercising such a disposition in making an utterance, but this is no special problem for the theory of names: we have the same problem with exercising the dispositions that constitute any skill: these are tricky problems in psychology (2006a: 210–220; Forthcoming-b). Despite what remains unknown, it should *go without saying* that there is a linguistic rule, established by convention, connecting, for example, 'Napoleon' to the famous French general, just as there is one connecting 'dog' to certain familiar pets and 'cat' to certain others. For, the alternative is that 'Napoleon', 'dog', and 'cat' are not part of our language. And if they are not, what is? We are *en route* to the batty conclusion that we don't have a language.[26]

We shoud try to avoid attributing such a view to Donnellan. And I think we can by taking the passages in Donnellan that have led Bianchi and Bonanini to their interpretation as better explained as reflections of Donnellan's insensitivity to the Gricean distinction between speaker-reference and semantic-reference. As Bianchi notes, Donnellan does not make the distinction in his discussion of names (##17).

Consider, for example, the claim by Bianchi and Bonanini that, according to Donnellan,

> (*) once someone has an individual in mind, in order to refer to it he or she may in principle use whatever name (or other expression) he or she likes. (2014: 193)

(*) is true if taken to be about speaker-reference *but not about semantic-reference*. This is demonstrated by cases where the speaker-referent is not the semantic-referent. Suppose that someone uses 'Napoleon' according to the convention and so semantically refers to the great general. Now *nearly always*, she will have that general in mind – the thought she means to express will be about him – and so will speaker-refer to him. But *not always*. Consider my example of the cynical journalist observing General Westmoreland at his desk during the Vietnam War and commenting metaphorically: "Napoleon is inventing his body count" (1996: 225–227; 2015b: 126). The journalist has Westmoreland in mind yet the semantic-referent of 'Napoleon' is the famous French general and *could not be* Westmoreland. For:

> A linguistic name token can have the conventional meaning, *M*, as a result of the speaker's participation in a convention only if *there is* a convention of using tokens of that physical type to mean *M*" (1996: 226–227).

So, it is not true that the journalist could "use whatever name (or other expression) he or she likes" to *semantically* refer to Westmoreland.

(1) *The novel uses* of names yield the first sort of case of a speaker-referent being different from a semantic-referent. (a) Metaphorical uses, like that of the journalist,

[26] Surely the neo-Donnellanians do not embrace this conclusion. Still, Almog et al. (2015) do leave me wondering about their answers to these crucial questions: What is a language? What is having one supposed to do for us and some other animals?

provide examples of this novelty. (b) Name introductions through use not dubbing – think of many nicknames and pseudonyms (1974: 199; 1981a: 58; 2015b: 114) – provide others. Novel uses demonstrate the crucial role of the speaker/semantic distinction in the theory of language. I said earlier that "the contents of thoughts *create*, *sustain*, and *change* the meanings of linguistic expressions". The contents do this in that they constitute speaker-meanings, and regularities in speaker-meanings create, sustain, and change semantic-meanings (2013d: 96–99; Forthcoming-b). *We need the distinction to explain the origin of linguistic meaning and language*. (2) The *confused use* of a name, like in the raking-the-leaves example (Sect. 19.2.2), provide another sort of difference: whereas 'Jones' lacks a determinate speaker-referent, its semantic-referent is Jones and *could not be* Smith because there is no linguistic rule linking it to Smith. (3) *Spoonerisms* provide another. Spooner was right when he finished a sermon, "When in my sermon I said 'Aristotle' I meant St. Paul": Aristotle was the semantic-referent, St. Paul, the speaker-referent (1981a: 139–146; 2015b: 126–127).

Bianchi and Bonanini are *clearly* right to attribute (*) to Donnellan if (*) is about speaker-reference but not, I would argue, if it is about semantic-reference. And there is another truth which they do *not* attribute, but which we should charitably suppose Donnellan would have embraced once sensitive to the speaker/semantic distinction: that *in virtue of having borrowed the semantic-reference of the name* a person can have an object in mind, can use the name to speaker-refer to the object, *and can use the name to semantically-refer* to it.[27] And that's how we all use 'Aristotle' to semantically-refer to the ancient philosopher. Of course, this truth presupposes a reference-borrowing theory of semantic-reference. And attributing that theory to Donnellan makes his view like Kripke's, just as people have commonly supposed it is. This attribution is surely preferable to the Bianchi-Bonanini one and, so far as I can see, none of the evidence Bianchi and Bonanini produce for their interpretation shows that Donnellan *denies* a reference-borrowing theory of semantic-reference. I conclude that Donnellan is not a neo-Donnellanian.

I emphasize that my proposal is not that we should take "Donnellan's considerations on proper names as concerning speaker's reference rather than semantic reference", a proposal that Bianchi and Bonanini reject forcefully: "it is indisputable that Donnellan's main critical target … is a *semantic* claim about proper names" (2014: 198). Indeed, I think that Donnellan's remarks, including those suggesting reference borrowing, should *mostly* be taken to be about semantic-reference. My point is that because he was insensitive to the distinction, his remarks should *sometimes* be taken to be about speaker-reference.

[27] Bianchi and Bonanini do attribute something like this to Donnellan in a footnote (2014: 194–195 n. 36) but, for reasons that escape me, deny that what they attribute is *Kripkean* reference borrowing. In any case, what I am suggesting that we should attribute to him seems to me to clearly fit their description of the Kripkean view: "any token of a proper name, except for the first, inherits its reference from preceding ones, to which it is historically connected.… [N]o further reference fixing is required" (2014: 195 n. 36).

One reason I'm in favor of bestowing this small bit of charity on Donnellan is that I'd like a similar bit to be bestowed on me. For, my early presentation of the causal theory (1972, 1974), though not my later ones (1981a, 2015b), were also insensitive to the speaker/semantic distinction.

It can be seen that I regard the speaker/semantic distinction as absolutely fundamental to semantics. Bianchi (see his forthcoming "Reference and Language") and the neo-Donnellanians (e.g. Capuano 2018; Pepp 2018; Almog et al. 2015) do not.[28] Is this the root of the problem in this dialectic?

So I see Donnellan's picture, charitably interpreted, as like Kripke's. I think that Bianchi is right to see some similarities between neo-Donnellanianism and my old view, most obviously the causal-perceptual view of reference fixing, but those similarities sit in very different theoretical frameworks.

Turn now to a comparison of my theory with Kripke's. I summed up my theory as follows in "Still Problematic" (a paper provoked by Bianchi!):

> *Speaker-Designation*: A designational name token speaker-designates an object if and only if all the designating-chains underlying the token are grounded in the object. (2015b: 125)

> *Conventional-Designation*: A designational name token conventionally-designates an object if and only if the speaker, in producing the token, is participating in a convention of speaker-designating that object, and no other object, with name tokens of that type. (2015b: 126)

Now, the social mechanism of reference borrowing plays a key role in my account of designating-chains. Reference-borrowing is often a crucial part of the determination of the speaker-reference of a name token and near enough always a crucial part of the determination of its semantic-reference. So my view seems to have what Bianchi requires for it to be Kripkean. Yet he says that this "final, 'official', formulation" of my theory "does not appear" to be Kripkean (##20). Why not?

Bianchi thinks not because of the role I give to the cognitive state of the speaker in determining reference, an idea he sees as coming from Donnellan not Kripke:

> there is no mention at all of having-in-mind, or related cognitive states or events, in the second lecture of *Naming and Necessity*…. [H]aving in mind does not play any deep explanatory role in Kripke's chain of communication picture. (##15)

Now we should start by setting aside my *introduction* of the cognitive state that I take to be constitutive of reference with talk of "having in mind". For, as I point out,

[28] The reason Almog et al. give for rejecting this distinction is that a convention cannot produce "a bond" between a name and its bearer (2015: 369–370, 377). Clearly I think it can and does: the "bond" is a conventional linguistic rule linking the name to its bearer, a rule established by the regular use of the name to speaker-refer to the bearer. I wonder if the neo-Donnellanians miss the importance of the speaker/semantic distinction for names because they assimilate names to simple demonstratives like 'this'. For, it is plausible to say, initially at least (but see later on the need to get away from this vague ordinary talk), that 'this' is governed by a conventional linguistic rule that makes it semantically refer to the object that the speaker has in mind (2004: 290–291); so it semantically refers to its speaker-referent. But a name is not a simple demonstrative and its semantic-reference is governed by a different rule.

and Bianchi notes (##16), this talk is but a "stepping stone" (1974: 202) which does not feature in my causal theory; see summary above. Still, Bianchi's discussion has made it clear to me that I should have been more careful in my handling of this stepping stone. I should not have claimed, early, to offer a causal "analysis" (1974: 202) of "having in mind" and, late, "an explanation – better, an explication – of this somewhat vague folk talk" in causal terms (2015b: 111). The folk talk arguably covers more than the causal relation we want. So my causal explanation was, in reality, an explanation of a "technical" notion of having-in-mind that is more restrictive than the folk notion.[29]

The importance of getting away from vague folk talk to a causal explanation is nicely illustrated by Pepp's response to Genoveva Martí's criticism of neo-Donnellanianism (2015).[30] Pepp's key thesis is "*Cognitive Priority*", "the metaphysical thesis that names in particular uses refer to things in virtue of speakers thinking of those things" (##2), in virtue of their "hav[ing] in mind to refer to" those things (##5). Martí offers a counterexample (2015: 80–81) to which Pepp responds. The counterexample is similar to the raking-the-leaves case and so, for convenience, I shall adapt Pepp's response to the counterexample as if it were to that case. Pepp thinks that even if she allows that "the only referent of the use [of 'Jones'] that is in accord with linguistic convention" is Jones, "this does not require rejecting Cognitive Priority". For, the reference to Jones is "still partially in virtue of the utterance being generated by the speaker's thinking of [Jones]" (##14) And she is surely right: in *some* sense, the speaker is "thinking of", "has in mind", Jones[31]; and these are the

[29] Bianchi claims that my "causal explanation of a thought's aboutness is remindful of the one offered by Donnellan and elaborated by the neo-Donnellanians" (##15). Indeed, the neo-Donnellanians' explanation seems like a version of my old one. But is this explanation *Donnellan's*? Some neo-Donnellanians certainly take their theory to be an elaboration of Donnellan's; see papers in Almog and Leonardi 2012, particularly Almog 2012: 177, 180–182. For them to be right about this, Donnellan must have offered an explanation that includes *both* a causal theory of aboutness borrowing *and a causal-perceptual theory of aboutness fixing*. Now I think we should see a causal theory of borrowing as, at least, implicit in Donnellan (and Kripke!), but what about the causal-perceptual theory of fixing? So far as I can see, neither Almog nor anyone else cites any evidence for the attribution of this theory to Donnellan. I found no such explanation in Donnellan's early papers, as I have noted (1981a: xi, 283–284, n. 12; 2015b: 111 n. 4): there seems to be no sign of a theory that explains fixing in terms of both causality and perception. As Julie Wulfemeyer has remarked recently: "The grounding cognitive relation went largely unexplained by Donnellan" (2017: 2). Finally, we should note Donnellan's defense of his theory against the charge of being "excessively vague": he compares his theory with "the causal theory of perception" (1974: 18–19). But, the vagueness of the causal-perceptual theory is not just *similar to* that of the causal theory of perception, it is *the very same*. (It is, in effect, the qua-problem for the theory of grounding; see Sect. 19.2.4.). If Donnellan held the causal-perceptual theory of fixing he surely would not have mentioned the causal theory of perception without noting that he held that theory of fixing!

[30] Also by Capuano's distinction between Donnellan's theory, "DPN", and Kripke's, "KPN" (2018: ##3).

[31] Martí prefers her counterexample to raking-the-leaves because she thinks that her analogue of "the speaker is 'thinking of', 'has in mind', Jones" is less plausible (2015: 82–83 n. 12; Martí has reversed the roles of Smith and Jones). Probably so, but it is still plausible enough, in my view (Sect. 19.2.2).

expressions that feature in Cognitive Priority. Yet we would surely all be drawn also to the idea that, in *some* sense, the speaker is "thinking of", "has in mind", Smith. In light of this, Pepp's claim that "the Conventional Stance rejects Cognitive Priority" (##4) is not clearly true. For, the Stance's view that reference to *x* is "determined by linguistic convention" may be quite consistent with that reference being determined by "thinking of *x*"; participating in the convention may involve "thinking of *x*", in some sense. We need to move beyond this folk talk to identify a real difference between Pepp and the Conventional Stance.[32]

I move beyond by telling a causal story in terms of two notions, *underlying* and *participating in a convention*:

> *Underlying* concerns the process of a speaker using the name to express a thought grounded in a certain object. *Participating in a convention* concerns the process of a speaker using the name because she has a disposition, dependent on the dispositions of others, to use it to express thoughts grounded in a certain object. Typically these two groundings are in the same object (2015b: 126)

Our common use of 'Aristotle' to refer to the great philosopher exemplifies what is typical. We express a thought grounded in him *by exercising our disposition* to express such thoughts using 'Aristotle'. The great philosopher is the speaker-referent because of his grounding relation to the thought expressed. He is the semantic-referent because of his grounding relation to the disposition exercised in expressing the thought.

In atypical cases, the speaker-referent and semantic-referent are different. (1)(a) In the journalist's novel metaphorical use of 'Napoleon', he expresses a thought grounded in one object, Westmoreland (speaker-referent), by intentionally exercising his disposition to express thoughts grounded in another object, Napoleon (semantic-referent). (b) In novel uses like the introduction-in-use of a nickname, the speaker expresses her thought grounded in an object (speaker-referent) without exercising *any* disposition to use that name to express thoughts grounded in an object (no semantic-referent). (2) In confused uses like the raking-the-leaves case, the speaker uses 'Jones' to express her thought grounded in two objects, Smith and Jones (partial speaker-referents),[33] by exercising her disposition to express thoughts grounded in just one, Jones (semantic-referent). (3) Canon Spooner uses 'Aristotle' to express a thought grounded in one object, St. Paul (speaker-referent), by accidentally exercising his disposition to express thoughts grounded in another object, Aristotle (semantic-referent).

So, the Conventional Stance, as I understand it, is committed to there being reference-determining processes of exercising such name-object dispositions, established by conventions. It is not committed to this process being free of cognitive activity that one might loosely describe as "thinking of" the object.

[32] A similar terminological point needs to be made, I argue (2014c), in response to the claim by John Hawthorne and David Manley (2012) that there is no causal "acquaintance" restriction on "reference" or "singular thought" (of the sort that the neo-Donnellanians and I urge).

[33] What I call "partial reference" Almog et al. seem to call "ambiguous reference" (2015: 371–373).

I certainly do not claim that the above view is Kripke's – his picture does not include the causal theory of groundings, for one – but why is it not Kripk*ean*?[34] Bianchi thinks not for two reasons. (1) He seems to think that giving *any* cognitive state a role in reference determination is unKripkean. But this is surely quite wrong. Notice that Kripke talks of intentions when discussing reference borrowing; see Sect. 19.2.1. And Kripke surely thinks that it is in virtue of a cognitive state of a speaker that her token of 'Aristotle' refers to the famous philosopher and not Onassis.[35] Indeed, to suppose that it is not partly in virtue of a speaker's cognitive state that her token has a certain reference is as implausible as supposing that we don't have a language. So why does Bianchi think that assigning a role to a cognitive state is unKripkean? (2) The answer seems to be that he thinks (##8) that to give a cognitive state this determining role is to *deny* a determining role to reference borrowing (perhaps Pepp would agree). This may be Bianchi's crucial mistake. For, the cognitive state plays its determining role *in virtue of being grounded in an object via reference borrowing*; see above.

In sum, Donnellan's and Kripke's pictures of reference determination, charitably understood, *both* include *both* a cognitive state of the speaker and reference borrowing. And I was right to present my theory as a development of these pictures in a naturalistic setting. But if I'm wrong on the interpretative issues, so be it. What we should all care about more is the theory not its origins. I claim that the theory is much more plausible than those of Bianchi or the neo-Donnellanians.

Bianchi has identified (##20–4) one respect in which my views may seem to be clearly at odds with Kripke's. As we have seen, I follow Grice in holding that conventional semantic-reference is explained in terms of speaker-reference. In contrast, consider this passage from Kripke, quoted by Bianchi (##22):

> we may tentatively define the speaker's referent of a designator to be that object which the speaker wishes to talk about, on a given occasion, and believes fulfils the conditions for being the semantic referent of the designator. (1979a: 15)

So, as Bianchi shows, it seems that Kripke must reject the Gricean priority of speaker-reference. But does he really? I thought not and so I asked him. He said that he does not. If not, what could be going on this quote? Kripke is offering an explanation of the speaker-referent of a designator that *presumes that the designator already has a semantic-referent*. Similarly, an explanation of the speaker-meaning of a metaphorical utterance would presume that the utterance already had a semantic-meaning. These explanations are no challenge to the Gricean idea that an utterance gets its semantic-meaning from speaker-meanings in the first place.

[34] Evans (1982: 69–79) labeled a causal theory of thought aboutness like mine "the Photograph Model" and claimed that it should not be attributed to Kripke. He rejects it in favor of "Russell's Principle" according to which to think about an object, "one must *know which* object is in question" (65). In response, I argue (1985: 225–227) that Kripke's ignorance and error arguments count not only against the description theory of names but also against Russell's Principle.

[35] On this see my 2015b: 110–111.

Finally, Bianchi mentions Kaplan's distinction (1989) between subjectivist and consumerist semantics. My theory of names is a bit of both. Insofar as the semantic reference of a person's use of a proper name is determined solely by her having borrowed the reference – for example, any contemporary person's use of 'Aristotle' referring to the famous philosopher – then the theory is consumerist: the name comes to her "prepackaged" with a semantic value.[36] Insofar as the semantic reference of a person's use is partly determined by her own groundings of the name[37] – for example, any associate of Aristotle's use of 'Aristotle' – then the theory is subjectivist: she is assigning a semantic value to it.

19.2.4 The Qua-Problem for Proper Names (Raatikainen, Reimer)

I wanted my theory of grounding to be "thoroughly causal" but there was a problem, "the qua-problem". The problem for names arises in two ways, as Sterelny and I (1999) pointed out. (1) In virtue of what is a name grounded in, say, my late cat Nana rather than a spatial or temporal part of her? For, whenever a grounder is in causal-perceptual contact with the cat, she is in perceptual contact with a spatial and temporal part of the cat. (2) Suppose that the would-be grounder is very wrong about what he is perceiving. It is not a cat but a mongoose, a robot, a bush, a shadow, or an illusion.

> At some point in this sequence, the grounder's error becomes so great that the attempted grounding fails, and hence uses of the name arising out of the attempt fail of reference. Yet there will always be *some* cause of the perceptual experience. In virtue of what is the name not grounded in that cause? (80)

Clearly there must be something about the grounder's mental state that requires the grounding to be in a cat, or something appropriately cat-like. But what is that something? I have struggled mightily with this problem, even hypothesizing that grounders must associate a "categorial" description, as Raatikainen notes (##12). With this hypothesis I moved from a causal to a "descriptive-causal" theory of grounding (1981a: 61–64; Devitt and Sterelny 1999: 79–80).

After presenting problem (1), Sterelny and I insisted that it was not to be airily dismissed on the ground that we just *do* designate "whole objects", for we do not always do so, as we illustrated; "there must be something about our practice which makes it the case that our names designate whole objects" (1999: 79).

[36] However, as Bianchi (##24) and I (2015b: 116 n. 17) note, Kaplan's consumerism involves backward-looking "deference". I want no part of that (Sect. 19.2.1).

[37] The semantic-reference of her use is almost certainly only *partly* thus determined because even though she has grounded the name herself, every time she hears and correctly understands another's use, she "borrows" it, thus reinforcing her ability to use it with that semantic reference (2015b: 117).

Marga Reimer has a different view. In "The *Qua*-Problem for Names (Dismissed)" she describes the qua-problem for names as a "pseudo problem [that] is easily dissolved" (##7):

> This problem can arguably be dismissed (*pace* Devitt and Sterelny) by appeal to a psychologically motivated default practice of naming whole objects rather than parts thereof…. The existence of a default practice of naming (only) whole objects would be easy enough to explain: such a practice has a clear *psychological motivation*. As natural language speakers, we have a strong practical interest in thinking about, talking about and (in some cases) beckoning or and otherwise addressing, whole objects (notably "persons, animals, places, or things") that are especially significant to us. (##5)

Reimer is persuasive that there is indeed "a *psychologically motivated* default practice" of naming such "whole objects" as persons, animals, and places that she mentions. This explains how we come to *be* naming them. But this is beside the point. For, the point is not about the *cause* of us naming them but about *what it is* to name them: "*In virtue of what* is what we do the practice of naming persons, animals, and places? *What makes it the case* that our practice is that? *What constitutes* its being the case that we mostly name those things?" Rather than naming Nana, we *could* have named her tail. And, as a matter a fact, people did name Sydney, which is a part of New South Wales, which is a part of Australia, … There must be processes that constitutes our naming Nana rather than her part and Sydney rather than what it is a part of (or Balmain which is part of it).

Reimer accepts, of course, that we often do name the parts of objects but she supposes that we do so only by *pretending* that they are "whole objects": they are "conceptualized as wholes by would-be grounders":

> Perhaps we can then regard *all* groundings of names as involving the practice of naming only whole objects, whether real or "pretend." In that case, such a practice would no longer really be a "default" practice; it would be an impossible to fault practice. (##6)

But to conceptualize something – Sydney, for example – as a "whole object" seems to amount to nothing more than treating it as nameable! And *anything* is nameable. Perhaps our use of 'whole object', initially in scare quotes, has misled here. For, every part of a "whole object", whether Australia's Sydney or Nana's tail is itself also a "whole object" that can be named. Reimer has not told us in virtue of what 'Sydney' names the city not Australia, 'Nana', the cat not her tail.

In the course of making this point in communication with Reimer, I suggested that we imagine different beings in a different environment who might have very different naming practices. Reimer addresses this (##11) but still, it seems to me, misses the point. These beings might be motivated to name time-slices of entities like Nana. Suppose that they did. Then there would have to be some processes in them, different from those in us, in virtue of which they are naming time-slices where we are naming cats.

So much for problem (1). What about problem (2)? A would-be grounder who believes she is perceiving a cat might be wrong: perhaps the cause of her experience is a mongoose, a robot, a bush, a shadow, or an illusion. How do we account for grounding failures in some of these circumstances? In virtue of what is the name not grounded in the cause of her experience *whatever it may be*? Reimer has a bold response: "I don't think that the question … is a worrisome one because the grounding will arguably *succeed* rather than fail in [these] cases" (##8). She then argues that it does succeed in those cases. This can't be right.

Let '*N*' be a schematic letter for a name. Now some substitution instances of the schema '*N* does not exist' are *true*. The custom is to call names in those instances "empty names". A theory of reference needs to *explain in virtue of what these names are empty and these instances of the schema are true*. Set aside non-paradigm "descriptive" names like 'Jack the Ripper' that have their reference fixed by a description. It is easy to explain the emptiness of one of those, as Reimer demonstrates in her discussion of 'Vulcan' (##13): nothing fits the description. But what about the paradigm names that, according to our theory, have their reference fixed by a causal grounding? The emptiness has to arise from a failure of such a grounding. So what does the failure consist in? That's the qua-problem (2).

Hallucinations provide the clearest case of emptiness. Reimer's discussion of one is puzzling:

> It's been one of those days and the speaker decides to help herself to a rather hefty portion of her husband's bourbon … and as a result has some extraordinarily vivid … dreams involving … a very large orange tabby cat…. It reminds her so much of her childhood pet Mieze that she decides to name the hallucinatory cat "Mieze." … [Later] she says …, "I wonder if I will see Mieze again tonight." A couple of weeks later … she says …, "I guess Mieze was just an incredibly vivid hallucination (##10)

Indeed it was. *And because it was we can say truly that Mieze does not exist and 'Mieze' is empty*. And the explanation that we must give of this is that *the grounding failed*. So Reimer's own discussion raises the qua-problem: In virtue of what does 'Mieze' not refer to whatever it was that caused her attempted grounding (the bourbon?)? Why does Reimer think otherwise? She emphasizes that the speaker has *successfully introduced the name* 'Mieze' and may continue using it and passing it on to others. Empty names can indeed be introduced and can have a long life. That poses a well-known problem.[38] But it is *a different problem.* Successfully introducing a name is much less demanding than successfully grounding it, as 'Mieze' illustrates. The causal theory needs an account of what makes the *grounding* of 'Mieze' unsuccessful. That's the qua-problem for names.

So what is the solution? As already noted, I moved to the idea that grounders must associate a "categorial" description. It's an ugly idea and I was reluctant to embrace it. One of its problems is that it lacks "psychological plausibility", as Reimer points out (##12). I now think, as she notes, that this view of a grounding is "far too intellectualized". I have taken to wondering whether this problem, like the reference-borrowing one above, "is more for psychology than philosophy" (2015b: 115 n. 15).

Insofar as philosophy has anything to contribute to the solution of the qua-problem, I think that we should look to teleosemantic explanations of the nonconceptual

38 I have attempted a solution (1981a: ch. 6).

content of *perceptions* (Devitt and Sterelny 1999: 162).[39] The basic idea is that the grounding of ‘Nana’ would be in a cat not a spatiotemporal part of her in virtue of the grounding involving a perception with the biological function of representing such an object not a spatiotemporal part of it.

19.2.5 Causal Descriptivism (Raatikainen, Jackson, Sterelny)

Raatikainen notes that Kripke did not present his ignorance and error argument as counting against all possible description theories. Raatikainen gives a critical summary of some that have emerged since Kripke. One that has been surprisingly popular is

> “causal descriptivism” favored by David Lewis (1984), Fred Kroon (1987), and Frank Jackson (1998) …. [S]peakers associate with a name ‘*N*’ a description of the form ‘The entity standing in relation *R* to my current use of the name “*N*”’, and this description determines the reference of “*N*”. The relation *R* here is drawn from the rival non-descriptivist (e.g. the causal-historical chain picture) theory of reference. (##23)[40]

The theory is, as Raatikainen says, “ingenious”. But it is also “fishy” (Devitt and Sterelny 1999: 61), not least because it is parasitic on the causal-historical theory.

In “Language from a Naturalistic Perspective”, Jackson asks “what’s the important difference between the two [theories]?” and puts his finger straight on an important one:

> Causal descriptivism is committed to the relevant causal and naming facts being pretty much common knowledge, whereas the causal theory proper holds that the naming and causal facts that secure the reference of a name are not common knowledge. (##7)

For competent speakers to know “the relevant causal and naming facts” is for them to “know the right ‘*R*’” to substitute in the above schematic form of causal descriptivism (Devitt and Sterelny 1999: 61).[41] Raatikainen objects, saying that this

> is *psychologically implausible*: it requires that every competent speaker must possess … the absolutely correct and complete theory of reference, and it is doubtful that anyone possesses such a theory. (##27)

This is not quite right (nor is Devitt and Sterelny 1999: 61). What causal descriptivism requires is that every competent speaker associates ‘*N*’ with the above “*R*-description” and hence, in effect, knows that *‘N’ stands in relation R to the referent, where R is the relation specified by the causal-historical theory*; this is “the relevant causal and naming fact”. So causal descriptivism does not require that

[39] On this, see an important recent book, Neander 2017.

[40] I first heard of causal descriptivism from Robert Nozick at Harvard in 1970 in response to my graduate student talk proposing a causal theory. Kripke also first heard the theory from Nozick; see 1980: 88 n. 38.

[41] Note that substituting ‘being the cause’ “is only the very vague and inadequate beginning of what is required” (Devitt and Sterelny 1999: 61).

speakers know *that very theory,* know that it is *in virtue of* such associations/knowledge that ‘*N*’ refers to its referent. Rather it requires that speakers have the knowledge on which the theory is built. Let us call this required knowledge, italicized above, “the knowledge-base” of causal descriptivism. And it is indeed “psychologically implausible” that every competent user of ‘*N*’ has that knowledge-base. Although causal descriptivism was designed to avoid ignorance and error problems it has a raised a new one.

19.2.5.1 Jackson

Jackson is unimpressed by this objection:

> This is, however, hard to believe. People write books and produce television shows on whether or not Helen of Troy or Jericho existed, and whether or not Shakespeare wrote the famous plays. These books and television shows report the results of historical research directed at the causal origins of the names in question and whether or not the origins are of the right kind to allow us to affirm that Jericho and Helen of Troy really existed, and that Shakespeare wrote, for example, *Macbeth*. The research that goes into these shows and books is carried out by historians, not philosophers of language. The shows and books get reviewed and assessed by people who are not philosophers, and the reviews are read and understood by the educated public at large. How could all this be possible if which facts are relevant to answering the questions were known only to a select group of philosophers? (##8)

Jackson is surely right that *many historians and other educated people* are quite good at identifying named entities like Helen of Troy. This is not to say, of course, that *any competent user* of *any* name knows how to identify its bearer. Nonetheless, Jackson does seem to think that she does:

> If you say enough about any particular possible world, speakers can say what, if anything, words like ‘water’, ‘London’, ‘quark’, and so on refer to in that possible world. (1998: 212).

And Jackson’s line of defense of causal descriptivism requires at least that. Yet note that knowing how to identify an entitity requires knowing not only what evidence counts, which is hard enough, but also knowing *when one has enough evidence*, when you *have* said enough “about any particular possible world”. It strikes me as a romantically optimistic view of the epistemic capacities of our species to suppose that we all have this ability.

But suppose, for the sake of argument, that we all did. The defense still has an interesting other problem that can be most persuasively presented by appealing to cognitive psychology. There is a familiar folk distinction between knowledge-*that* and knowledge-*how*. Psychologists and cognitive ethologists take this distinction to be much the same as their own very important one between “declarative” and “procedural” knowledge. Thus, John Anderson, a leading cognitive psychologist, writes:

> The distinction between *knowing that* and *knowing how* is fundamental to modern cognitive psychology. In the former, what is known is called *declarative knowledge*; in the latter, what is known is called *procedural knowledge*. (1980: 223)

Psychologists describe the distinction, rather inadequately, along the following lines: where declarative knowledge is explicit, accessible to consciousness, and conceptual, procedural knowledge is implicit, inaccessible to consciousness, and subconceptual. Finally, *skills* are procedural knowledge not declarative knowledge.[42]

Now knowing how to identify an entity *x* named '*N*' is a *cognitive* skill, a piece of procedural knowledge. It consists in knowing how to discover, and make effective use of, alleged information about *x*, information about the properties of *x* that will be presented in claims of the form, '*N* is *F*'; and it consists, as just noted, in knowing when enough is enough. But this procedural knowledge is not what Jackson's defense requires. It requires common knowledge of "the relevant causal and naming facts"; of "the knowledge-base" of causal descriptivism; of the fact that *'N' stands in relation R to the referent, where R is the relation specified by the causal-historical theory*. This knowledge-base is, of course, a paradigm of declarative knowledge. It is quite different from the procedural knowledge that Jackson supposes we all have, both intuitively and according to the consensus in psychology: *procedural knowledge is not declarative knowledge*.

But mightn't procedural knowledge *cause* declarative knowledge? Thus, one might observe the exercise of one's skill at, say, riding a bicycle, typing, or thinking, and abstract the principles of a good performance. Similarly, one might observe the exercise of one's skill at identifying *x*, the referent of '*N*', and abstract the principles of reference. And one surely *might* but, importantly, one *typically doesn't*. Indeed, most of us *typically can't*; we are just not clever enough. And even where we are clever enough to discover the nature of a skill, that does not constitute the skill. Our epistemic success is one thing, the metaphysics, another.

In general, knowledge-how is very different from knowledge-that.[43] In particular, knowing how to identify an entity is very different from having the knowledge-base required by the causal descriptivist theory of the entity's name. Even if all competent users of a name '*N*' knew how to identify its bearer by using descriptions of the bearer, which they surely don't, we have no reason to believe that the *R*-description, 'the entity standing in relation *R* to my current use of '*N*'', *with the "correct" substitution for 'R'*, is even among the descriptions used by *anyone* for this purpose. Raatikainen rightly says that it is "doubtful that anyone possesses" the right theory of reference, even philosophers (##27); it's just too difficult.[44] And it is *really* "hard to believe" that the folk have the knowledge-base of causal descriptivism.

[42] My claims about cognitive psychology draw on my examination of this and its relation to language in *Ignorance of Language* (2006a: 210–220); those about cognitive ethology, on "Methodology and the Nature of Knowing How" (2011c: 213–215).

[43] Stanley and Williamson (2001) disagree. I have argued that they are wrong (2011c).

[44] Jackson says that causal theorists "must of course allow that some people are aware of the relevant facts – namely, they themselves" (##7). Well, we do keep our spirits up by thinking that we are aware of *some* of the relevant facts but it would be *wishful* thinking to suppose that we were aware of them *all*; note the acknowledged incompleteness of the theories of reference borrowing and grounding discussed in Sects. 19.2.1, 19.2.2 and 19.2.4.

Mean aside. If that knowledge-base was really "common knowledge", it is curious that there was hardly even a glimmer of causal descriptivism in the theories proposed in the pre-Kripkean era.

Jackson thinks that not only are "the relevant facts (about naming procedures and information preserving causal chains containing names) … widely known, but [they] had better be widely known. Only that way can proper names play their important informational role" (##8). But all that is required for 'Hillary' to play its role of conveying information about Hillary Clinton is that people participate in the convention of using 'Hillary' to convey thoughts about Hillary (Sect. 19.1.2). This again is a piece of knowledge-how. It does not require people to have causal descriptivism's knowledge-base for 'Hillary'. Indeed, in my view, it does not require them to have any semantic propositional knowledge at all (2006a).

Jackson has an over-intellectualized view of linguistic competence. This underlies another step in his defense of causal descriptivism. According to that theory, the reference-determining description that users of '*N*' associate with it, its *R*-description, is metalinguistic; see above. Jackson defends this metalinguistic commitment as follows. Suppose that you wanted to discover whether Mary Smith works in this building. Jackson claims that

> it is impossible to show that 'Mary Smith works in this building' is true without having recourse to information about words, and in particular the words 'Mary Smith'. You need to find them on an office door, or in the phone directory of the building, or to note the response of someone who works in the building to a question containing 'Mary Smith' – as it might be, the question "Does Mary Smith work in this building?" (##10)

This seems quite wrong to me. Thus, to learn from that worker's testimony, you simply need to understand what she says. This requires a skill at using words, a piece of knowledge-how. There is no reason to think that it requires any "information about words", a piece of knowledge-that; or so I have argued (2006a).

In sum, I think that Jackson has not rebutted the ignorance and error objection to causal descriptivism. Competent speakers do not all know how to identify the bearers of the names they use. Even if they did they would not thereby have the knowledge-base required by causal descriptivism.

Sterelny and I had another, deeper, objection (1999: 61). Raatikainen puts it nicely:

> [causal descriptivism] is *parasitic and redundant:* if it is true, it admits that a name stands in a causal-historical relationship, *R,* to its bearer; *R* alone is sufficient to explain reference, and further description involving *R* is redundant. (##23)[45]

The point is that, according to causal descriptivism, reference is determined by the *speakers' association of the R-description with 'N'* and that association must be redundant. We can come at this redundancy from the other direction. Suppose causal theorists finally discover that '*N*' refers to *x* in virtue of *x* standing in a certain causal

[45] Note that this is not a model for a suspiciously quick refutation of *any* description theory. Causal descriptivism's redundancy comes from its peculiar feature of *having a reference-determining description that specifies the reference-determining relation.*

relation to our use of '*N*'. Obviously those theorists will describe that relation in presenting their theory! Hence they will associate that description with '*N*'. But it won't be their association of the description, let alone ordinary speakers' association of it, that determines reference: reference is determined by the relation described.

Jackson does not address this objection and Sterelny, in "a shocking betrayal", seems to have forgotten it in his present flirtation with Jacksonian descriptivism. For Sterelny, in "Michael Devitt, Cultural Evolution and the Division of Linguistic Labour" claims to have "bought into something like Jackson's causal descriptivism" (##3). I am hoping that what follows will lead him to repent.

19.2.5.2 Sterelny

Sterelny addresses two questions. The first concerns what we "need to know or do to be able to successfully launch a new term into the linguistic community" (##2). In effect, this is the issue of reference fixing (see Sects. 19.2.2, 19.2.3 and 19.2.4) The second concerns "how much information needs to flow down the links for a novice to acquire the capacity to use the term" (##2). In effect, this is the issue of reference borrowing (Sects. 19.2.1, 19.2.3). Sterelny takes me to have "staked out a minimalist position with regard to both questions". In contrast, he now leans toward the alternative view "that both launching a term, and acquiring through conversational interaction with others using it, is more informationally demanding" (##2). And his concern is not just with proper names but also with kind terms, particularly "natural kind" terms.

How minimal is the position that I have "staked out"? Consider names first. When it comes to their reference borrowing, my position is indeed minimal in its demands on "the central processor". However, I anticipate that psycholinguists will reveal that lots of subdoxastic processing is necessary; see the discussion in Sect. 19.2.1. When it comes to reference fixing, as noted in Sect. 19.2.4, I once embraced, reluctantly, an informational demand. So I was not then a minimalist. However, as also noted, I have backed away from that demand. What about "natural kind" terms? My position on the two questions is much the same for them as for names (Devitt and Sterelny 1999: 88–93). When it comes to *other* terms however, I, in the company of Sterelny himself, have attempted to describe the logical space for positions *but staked out none* (1999: 93–101). I think we will need to start testing usage to make progress here (Sect. 19.4.4).

So, it would be fair to say that minimalism is my default position. Still, aside from proper names and (some?) "natural kind" terms, I think it is an open question whether information about *the referent* must go into fixing or borrowing. But I do insist that nothing like causal descriptivism's knowledge-base, nor any other *semantic* information, is *constitutive* of the reference fixing or borrowing of any term. Sterelny's interesting ideas about the evolution of language here and elsewhere (2016) do not support anything like Jacksonian descriptivism.

Sterelny does seem to think otherwise: “In using a referential term, speakers are aware of, and buy into, the causal networks that link their uses of a term to its bearer” (##3). Why does he think so?

> [T]he flexibility of semantic competence was hard to reconcile with Michael’s automatic, quasi-reflexlike picture of the introduction and transmission of a term from one speaker to the next…. I can coin a new name – for the recently arrived dog – and smoothly integrate it into my language. If it is apt, it is likely to catch on and become part of the local dialect. This strongly suggests that agents represent the referential, semantic properties of words – we notice them …. Moreover, this is a central aspect of their semantic competence, not a peripheral one…. Our capacity to use names … depends not just on the existence of these sociolinguistic networks. It depends as well on our recognition of their existence, and on our intention to use a name in a further extension of the network. (##3)

It is not clear to me why Sterelny thinks that the “automatic, quasi-reflexlike picture”, or indeed *any* picture, of linguistic competence is “hard to reconcile” with the flexibility in reference fixing that he rightly emphasizes: that “we can expand our lexicon, at will, at need, on the fly” (##3). But suppose the reconciliation is indeed impossible. Why would this “strongly suggest” that the picture has to be enriched by *semantic* information rather than the sort of information about the referent that I have just agreed may well be required for some terms?

Set that aside. I want to focus on two further matters that, on the one hand, accommodate Sterelny’s idea that language acquisition is informationally demanding but, on the other hand, do not go down the Jacksonian path. The first matter is on reference fixing, the second, on reference borrowing.

(I) I agree that the capacity to coin new words that Sterelny describes is a “central aspect” of our (innate) competence *to learn natural languages*. So this capacity partly *causes* lexical expansion, partly *causes* our competence *in any language we do learn*. Further, this capacity surely requires a lot of cognitive sophistication, as Sterelny insists. In particular, it surely requires lots of thoughts about other minds. I certainly think that we should *entertain*, though *be reluctant to embrace*, the idea that the capacity requires thinking some semantic thoughts.[46] Indeed, I once made a “tentative” proposal of this sort: “Some sort of semantic theorizing is required to gain linguistic competence, including some elementary theorizing about the point of language” (1981a: 109). It seems to me that Sterelny’s idea that language acquisition is informationally demanding is most plausible if construed as a proposal of that sort: semantic thoughts are part of the *cause* of our capacity to coin new words. I have two points to make about this proposal.

First, this causal connection does not make the semantic thoughts *constitutive* of that competence in the coined words (1981a: 109). So such proposals are not even in the same game as causal descriptivism. A look at psychology helps again. Psychology distinguishes two sorts of learning, *explicit learning*, which is a

[46] The case of the prairie dogs alone should give pause. Prairie dogs have a system of barks that convey information about which sort of predator is threatening. When an experimenter used a plywood model to simulate a new sort of predator, the prairie dogs introduced a new bark (Slobodchikoff 2002).

"top-down" process, and *implicit learning*, which is a bottom-up process. Set aside implicit learning for a moment. Explicit learning *starts from* declarative knowledge. Consider learning to change gears in a stick-shift car by starting with instructions like: "First, take your foot off the accelerator, then disengage the clutch." That's one example of explicit learning. The proposal that semantic thoughts partly cause our capacity to coin new words is another example. And the important point is that in explicit learning, declarative knowledge, like the instructions and the semantic thoughts, plays a role in *bringing about* the skill, but *does not constitute* it and may even disappear after the skill is learnt. The psychological consensus is that the skill is procedural knowledge, knowledge-how.

Second, supposing that we must have *some* semantic thoughts for language acquisition may have some plausibility, but supposing that we must have something like *the knowledge-base for causal descriptivism*, as Sterelny is suggesting, strikes me as very implausible.

My first point rests on a crucial distinction between what causes competence and what constitutes it. Sterelny's discussion may support the idea that some true semantic thoughts are included in the *cause* of competence. I'm dubious, but if the discussion does support this, then causal descriptivism may not be *as* open to the ignorance and error objection as we once claimed. But the discussion does not support the idea that true semantic thoughts, let alone the knowledge-base of causal descriptivism, *constitute* the competence. Of course, as a causal theorist, I must allow that causes of some sort *could be* constitutive, but I don't think that we have been given *any reason to believe* that semantic thoughts are constitutive of linguistic competence. My second point is that, even if semantic thoughts are part of the cause of the competence, it is very implausible that the knowledge-base is. So, causal descriptivism still has an ignorance and error problem.

(II) That point is about reference fixing. Sterelny thinks that causal descriptivism may get support from a consideration of reference borrowing:

> Perhaps information about the cognitive demands of cumulative cultural learning in general can give us some insight into the particular case of what an agent must understand in order to be part of reference borrowing networks, to be part of the division of linguistic labour. (##7)

Getting this support is difficult because "there is no received view on the cognitive prerequisites of cumulative cultural learning in general" (##7). Sterelny's interesting discussion explores two schools of thought. First, there is the "relatively undemanding 'Californian' view". This view of cultural learning seems to fit the psychologists' idea of *implicit learning*, a "bottom-up" learning that takes place "largely without awareness of either the process or the products of learning" (Reber 2003: 486). Clearly, there is no support for Jackson there, but Sterelny does see some in the alternative "Parisian" view

> which sees cultural learning as more cognitively demanding …. [T]he process is highly selective, and what agents do with the representation that they do pick up depends on what they already know or believe; their intentions; and on how their mind is tuned. (##8)

Sterelny thinks this view "obviously resonates with Jackson's reflective conception of the sociolinguistic networks that sustain the division of linguistic labour" (##9). I don't think so.

The "Parisian" view seems to fit the psychologists' idea of *explicit learning*, a "top-down" process that I have described above. And the important point now, as in (I) above, is that explicit learning and the Parisian view are theories of how we achieve a skill, about its *cause*, they are not theories of what *constitutes* that skill. So, no resonation with Jackson.

Sterelny argues that reference borrowing is at least partly Parisian:

> the flexibility of use that I took to support Jackson's view that the division of linguistic labour depended on downstream users recognising the networks of which they are a part may be a relatively recent feature of language, perhaps most marked in large world languages. That said, on this view the languages (or protolanguages) without a word for word, or other ways of explicitly thinking about words and reference, will also be languages without much (or with very minimal) division of linguistic labour. (##11)

So for speakers to borrow reference they must have semantic thoughts about words and reference. This is an interesting idea and, taken as a view of the cause of the borrowing, it may be right. Still, I think there are reasons to doubt it. When I looked hard at the matter, admittedly a decade ago, I concluded that "language learning seems to be a paradigm of implicit learning" (2006a: 219). I'm still inclined to that conclusion. If it is right, declarative knowledge, whether semantic or not, plays no role. Still, we should all agree, it is early days in understanding these processes.

But, to repeat, even if language learning is explicit, with declarative knowledge playing a causal role, that does not support Jackson. For, causal descriptivism is a theory of what constitutes reference. So Sterelny must have in mind that these semantic thoughts are constitutive. Yet, in the context of discussing reference borrowing in an earlier paper, he remarks:

> While it is hard to specify exactly what users understand about their language in using it, and, certainly, they have nothing like an explicit commitment to any semantic theory, referential competence is not reflexlike, or completely isolated from reflective understanding. (2016: 273)

It is not clear to me that this cautious statement is at odds with my "minimalist" view, described in Sect. 19.2.1, and hence is a very long way from causal descriptivism.

In sum, Sterelny leans toward an informationally demanding view of language learning and has some interesting thoughts in support. I am dubious that this learning demands semantic knowledge let alone anything close to the knowledge-base of causal descriptivism. But that is all about the cause of a person's competence with a name. My main point is that there is no case here for an informationally demanding view of the competence itself. In particular, I see no case for the view that the knowledge-base is constitutive of this competence. So the considerations he adduces should not take him so close to causal descriptivism.

Finally, I must remind Sterelny of our redundancy objection to that theory. This objection, arising out of the theory's parasitism, still strikes me as devastating. In light of the present discussion, we can expand on the objection as follows. Suppose, *per impossibile*, that a person knew that it was in virtue of '*N*' standing in a certain causal relation to the referent that '*N*' refers to the referent; i.e., she knew the correct causal theory. What role could that knowledge play? Not the role of *constituting* the reference of '*N*': see the redundancy objection. Yet that piece of declarative knowledge could *cause* the person to establish that reference-determining causal relation between '*N*' and an entity. So, that knowledge's role would be analogous to that of the instruction in learning to change gears in a stick-shift car.

19.3 Theory of Meaning

19.3.1 Direct Reference (Braun, Horwich)

David Braun and I have been arguing about direct reference (DR) for 25 years. Braun believes in it, I don't (1989a, 1996, 2012d). DR is more often declaimed than argued. I have always admired Braun's conscientious and ingenious attempts to argue for this unpromising doctrine, illustrated again in "Still for Direct Reference".

Braun's paper starts with nice accounts of DR and of my contrary view (for which I thank him). I shall presume these accounts in what follows. Braun goes on to give a long, thoughtful, and intricate response to my criticisms of DR. It would be a daunting task to respond to this in the detail it deserves. Sadly, I cannot attempt anything close to that here. I apologize for the resulting density of what follows.

DR, as we are understanding it,[47] is a resurrection of the "Millian" theory according to which the "meaning" ("semantic value", "semantic content", etc.) of a proper name is simply its bearer or, as I prefer to say, its property of referring to its bearer. So, it faces the familiar, and apparently overwhelming, problems that led Frege and Russell to abandon Millianism long ago in favor of description theories. Despite this, DR is commonly taken, by friend and foe alike, to be the theory of meaning required by the Kripkean revolution in the theory of reference. How did this widespread opinion come about? It's a curious history that I have recently tried to tell (2015b: 129–130).

I am a maverick among the revolutionaries. For many years I did not entertain DR as even a *candidate* theory. And, from the beginning (1974: 204), I presumed that the meaning of a name was to be found in the causal network that determined its reference, a network of groundings and reference borrowings (Sects. 19.2.1, 19.2.3 and 19.2.4): Frege was right to think that a name's meaning was its mode of reference but wrong to think that this mode was constituted by associated descriptions. We can safely say that a key reason for the success of the implausible DR is that the idea that a name's meaning is its causal mode of reference strikes people as

[47] "Direct reference" is often understood in other ways (1989a: 206–212; 1996: 170n).

too shocking to be taken seriously (2015b: 130–132). Still, I argue that, shocking as it may be to the tradition, it is right (2001).

However, it is important to note, I came to allow – in *Coming to Our Senses* (1996) – that a name did not have just this one causal-mode meaning.[48] The meanings we should posit are properties that play certain causal roles in the natural world. In particular, it is in virtue of their meanings that token thoughts and utterances cause behavior and guide us to reality. When it comes to guiding us to reality, a name token can mostly play its causal role simply in virtue of its property of referring to its bearer, simply in virtue of its DR meaning. But, when it comes to causing behavior, we need to posit a finer grained meaning, its property of referring to its bearer *in a certain way*, a mode of referring. And we should contemplate that the explanation of one piece of behavior may require a finer grained mode than the explanation of another.

So, the main objection to DR is what it omits from meaning, not what it includes. Interestingly, Kaplan (2012), a founding father of DR, has recently said as much![49] "The New Kaplan" urges, for reasons similar to mine, a semantic place for nondescriptive ways of referring that are also similar to mine. Yet he emphasizes that he does not want to deny a place to "Millian" meanings: "I only wish to resist the term 'semantics' being hijacked for one kind of content" (2012: 168 n. 28). This seems dead right to me. There is no sound theoretical basis for DR's insistence that there is just the one kind of meaning that DR favors.

I turn now to Braun's defense of DR.

(I) Braun sees an important difference between us over attitude ascriptions:

> an expression is *Shakespearean* iff substitution of co-referring proper names in that expression preserves the extension of that expression…. Direct-reference theory entails that attitude ascriptions and 'that'-clauses are Shakespearean, on all of their readings. (##4)

On my view, in contrast, these ascriptions are not Shakespearean on some readings. Why is this difference important? For me, not primarily because of what it shows about the meanings of these attitude ascriptions,[50] but because of what it shows about the nature of the properties *ascribed* by these ascriptions (1996: 70–81, 140–150). For, these ascriptions ascribe properties that seem to play the causal roles

[48] Braun notes that DR is concerned with "the *conventional* meanings of linguistic expressions" (#16) and is worried that my concerns may be elsewhere so that we are talking past each other. I don't think we are. Now it is true, as he notes, that my focus in *Coming* is on the meanings of thoughts rather than of utterances but I am also concerned with utterances. I later describe that concern as follows:

Paul Grice (1989) has drawn our attention to the distinction between the conventional meaning of an utterance and its speaker meaning…. In doing semantics, we are primarily interested in *conventional meanings on occasions of utterance* because those are the meanings that are the main routes to speaker meanings and hence thought meanings…. So the basic task for utterances is to explain the nature of properties that utterances have largely by convention and that play semantic roles. (2012d: 65; see also p. 79)

[49] My telling of the history of DR includes quite a lot on Kaplan, of course (2015b: 129–130).

[50] Although I do discuss these meanings, of course (1996: 82–84, 196–218).

of meanings and so *provide evidence* of the nature of meanings, evidence that meanings are those ascribed properties. Thus, whereas

(1) Ralph believes that Ortcutt is a spy,

construed transparently,[51] ascribes the property of referring to Ortcutt, *construed opaquely*, it ascribes the property of referring to Ortcutt under the mode of ‘Ortcutt’. I take this as evidence, *although not the only evidence* (1996: 150–154), that the latter property is part of a meaning (“content”) of the belief in question and hence of any utterance of ‘Ortcutt is a spy’ that expresses that belief. For, opaque ascriptions serve our purposes of explaining and predicting behavior:

> when we articulate the generalizations in virtue of which behavior is contingent upon mental states, it is typically an opaque construal of the mental state attributions that does the work. (Fodor 1980: 66)

Clearly, the opaquely construed (1) could not then be Shakespearean: substituting another name for Ortcutt, say ‘Bernard’, would change what it ascribes.

Now, some puzzle cases required adding a few bells and whistles to this theory. One puzzle that Braun focuses on is Kripke’s Pierre (1979b: 254–265). This case led me to say that in

(2) Pierre believes that London is pretty,

opaquely construed, and ascribed because Pierre is ready to assert “Londres est jolie”, ‘London’ ascribes “the ‘disjunctive’ mode of *‘London’ or ‘Londres’* rather than a finer-grained meaning involving one of the disjuncts” (1996: 233).[52] There is nothing special about the French version of ‘London’ and so I added a footnote to the mention of ‘Londres’ saying: “Or, indeed, the ‘translation’ of ‘London’ in any other language” (233 n. 81). This is the basis for Braun’s first criticism.

> Devitt can resist the Shakespearean conclusion only if ‘Londres’ *is* a translation of ‘London’, but ‘Bernard’ is *not* a translation of ‘Ortcutt’ (and ‘Ortcutt’ is not a translation of ‘Bernard’). Is there any reasonable translation relation that would have this consequence? (##28)

He argues ingeniously that there is not: I must settle for translation as co-reference and accept Shakespeareanism (##34). If this were so it would seriously undermine my case for meanings as causal modes of reference. Braun puts a lot of weight on this argument (##66).

(a) First, Braun takes my scare-quoted use of ‘translation’ in a footnote far too literally. My idea is simply that the opaque (2) ascribes a mode of referring to London by the causal network for ‘London’ *or by some closely related network*

[51] An expression is transparent iff substitution of co-referring singular terms in that expression preserves the extension of that expression. So where the definition of ‘Shakespearean’ talks of the substitution of proper names, that of ‘transparent’ talks of the substitution of singular terms.

[52] ‘London’ does not do so in all opaque ascriptions, however; consider “Pierre believes that London is not Londres” (235).

like that for 'Londres'. Just how closely related? It is hard to say: there is some indeterminacy here, as almost everywhere. Still, there are clear cases of *not being close enough*: the network for 'Mark Twain' is not close enough to that for 'Samuel Clemens'. Similarly, it is hard to say how close to a full head of hair a person has to have to be hirsute but Yul Brynner is clearly not close enough. And Ralph having a belief under the mode of 'Bernard' cannot be enough to make the opaque (1) true. So I don't have to accept Shakespeareanism.

These claims rest largely on intuitions about the truth values of attitude ascriptions. There is a more important consideration that is not about such "second-level" utterances.

(b) The Pierre puzzle does not undermine the argument, alluded to above, about the meanings that cause behavior (1996: 140–154). It follows from that argument that part of the meaning we have to attribute to Pierre's "first-level" utterance of "Londres est jolie", and to its underlying belief, *in order to fully explain Pierre's behavior* is the property of referring to London under the mode of 'Londres'; similarly, to his utterance of "London is not pretty", the property of referring to London under the mode of 'London'. Those are the meanings that fully explain Pierre's behavior.

Turn for a moment to Horwich. In considering my theory, he thinks that "one might well baulk at the idea that different primitive terms *never* have (indeed *cannot* have) the same meaning. What about 'London' and 'Londres'?" (##9) As can be seen, I think that the causal roles of a name justify our ascribing more than one meaning to it, some of which can be shared by another name. Thus 'London' and 'Londres' *do* share the following two meanings: the property of referring to London; the property of referring to London under the disjunctive mode of 'London' or 'Londres', or … But they *do not* share the property of referring to London under the mode of 'Londres'; only 'Londres' has that meaning. And its having that meaning is important in explaining Pierre's behavior.[53]

The Pierre puzzle does not show that meanings are not causal modes. Rather, it shows that that our ordinary attitude ascriptions are *not always* perfect ways to attribute the causally significant meanings: the disjunctive meaning that (2) ascribes "is not fine-grained enough to explain [Pierre's] behavior" (234). I had earlier found a similar imperfection in discussing demonstratives: Richard's puzzle (1983) shows that "the folk do not have standard forms of ascription that distinguish the meanings" of demonstrative beliefs that need to be distinguished to explain behavior in Richard's ingenious story (221).[54] And it is not hard to see why the folk do not have standard forms that will do the required explanatory job in these situations: the situations are unusual (231–234).

[53] Horwich (##9) also raises Kripke's case of Peter and 'Paderewski'. I have discussed this difficult problem (1996: 236–240).

[54] I contemplate positing "hidden indexicals" to avoid these charges of imperfection (221–222, 234–235).

Braun is not the first enthusiast for DR to get comfort from puzzles about attitude ascriptions. I responded to some earlier examples as follows:

> The conclusion that the direct-reference philosophers draw from the puzzles … are "in the wrong direction". The puzzles do not supply evidence that meanings are coarse grained but rather supply evidence that some meanings are *more* fine grained than those ascribed by standard attitude ascriptions. The puzzles show that we sometimes need to ascribe these very fine-grained meanings to serve the semantic purpose of explaining behavior. (243).

(II) Braun turns next to arguments against DR, particularly my versions of the Identity Problem and the Opacity Problem (173–176). After presenting these versions, he rightly says that they

> do not show direct-reference theory is false unless they also show either (a) that 'Twain is Twain' and 'Twain is Clemens' differ in conventional meaning or (b) that 'Alice believes that Twain is Twain' and 'Alice believes that Twain is Clemens' differ in conventional meaning. (##42).

He continues: "But Devitt's arguments do not show this." Well, they were certainly intended to (2012d: 79)! Let me make that explicit by adding 'conventionally' to a recent statement of the argument:

> suppose Abigail has a neighbor called "Samuel Clemens" and reads books by someone called "Mark Twain". On hearing someone report, "Mark Twain is at the Town Hall", she rushes there, saying, "I'd love to meet Mark Twain". It makes all the difference to our explanation of Abigail's behavior that she said this not "I'd love to meet Samuel Clemens". For, the former [conventionally] expresses a thought that causes her to rush to the Town Hall whereas the latter [conventionally] expresses one that might not have (because she may not know that Mark Twain is Samuel Clemens). It matters to the explanation of her behavior that her utterance [conventionally] refers to Twain/Clemens *under the mode of 'Mark Twain'*. (2015b: 133)

What could be the basis for denying that the mode of reference of 'Mark Twain' that is crucial to the name's causal role here is a *conventional* meaning of the name? At this point Braun follows the DR tradition of clutching at the straw of pragmatics (##42–43). But there is no principled basis for treating this phenomenon pragmatically; or so I have argued (1989a, 1996, 2012d). The use of the name in this story exemplifies the *undeniably regular* practice of using a name to convey a message partly constituted by the name's mode of reference. This regularity is best explained by positing a convention of using a name to convey such parts of messages; indeed, there is no other plausible explanation. So we should posit a convention (2013d: 107). So the mode of reference is a conventional meaning of the name.

(III) Appeals to pragmatics loom large in Braun's intricate defense of DR against the charge that DR cannot account for the causal role of utterances (##46–65). Indeed, that defense ultimately rests on exporting the troubling phenomena to pragmatics. This exportation strategy is demonstrated in two crucial, and characteristically honest, admissions at the end of Braun's discussion. The first concerns attitude ascriptions:

> I admit that attitude ascriptions that ascribe modes of reference provide more explanatory information about an agent's behavior than do (direct-reference-style) attitude ascriptions that do not. But I deny that modes of reference are among the meanings that attitude

> ascriptions ascribe to thoughts *in virtue of the conventional meanings of those ascriptions*…. [The] utterances of attitude ascriptions can *pragmatically convey* information about modes of references. (##63).

But the attitude ascriptions that provide that "more explanatory information" are not *occasional* occurrences, they are *regular* ones: attitude ascriptions with names are *standardly* used in a non-Shakespearean way. I stand by the charge that there is no principled basis for denying that this usage exemplifies a convention.

Braun likes to defend DR by arguing that attitude ascriptions, "second-level" utterances, are Shakespearean. I agreed that this issue is important because the non-Shakespearean nature of these ascriptions provides evidence that modes of reference are meanings of names. But it is much more important to look at the causal role of names in "first-level" utterances. Braun's second admission is helpful on this:

> I admit that modes of reference are theoretically interesting, in the following ways: modes of reference are real properties, some parts of thinking-events have those properties, they are causally relevant to behavior, and mentioning them (e.g., in technical ascriptions) can add explanatory information to explanations of behavior. But I also maintain that there is a principled and theoretically interesting distinction between conventional and non-conventional meaning. (##65)

I agree with every bit of this, of course. Now think of its implications. Consider one of those "thinking-events", Betty's belief that Twain smokes (one of Braun's example). Braun agrees that this belief plays its role in causing behavior partly in virtue of its property of referring to Twain under the mode of 'Twain'. So he should agree that this property is part of the *meaning* (content) of that belief; playing that causal role makes it a meaning. Let us call that thought meaning, partly constituted by the mode of 'Twain', "*M*". Obviously, Betty can express that belief by attempting to convey *M* to an audience, just as she can express any belief by attempting to convey its meaning. Grice (1989) and many others have emphasized that she might attempt this in many ways; for example, in certain circumstances, she might attempt it with the words, "Twain has a death wish". Whatever words Betty chooses for that purpose will "speaker mean" *M*, for the meaning of the thought she expresses determines what she *means by* her words. However, as an English speaker, Betty is *very likely* to use the sentence, "Twain smokes", for that purpose. She is very likely to use that sentence because it is the *conventional* English way of expressing a belief with meaning *M*. So the utterance of "Twain smokes" not only speaker means *M*, as does the utterance of "Twain has a death wish", but it also conventionally means *M*. So referring to Twain by the mode of 'Twain' is a conventional meaning of 'Twain'.

Where might Braun get off this train? He agrees that a thought's mode of referring is causally relevant to behavior and so is theoretically significant. (i) But he might resist the idea that this makes that property a meaning. What then does make a property a meaning? One wonders whether such resistance would be "merely verbal". (DR has often led me to such wonderings: 1989a, 1996, 2012d.) (ii) He is not likely to claim that whereas 'Twain smokes' is not the conventionally expression of *M*, some other some form of words is. (iii) We are left with this option: there is *no* conventional expression of *M*. This would be weird. Apart from the evidence that

'Twain smokes' *is* such a conventional expression, it would be a truly remarkable failure of our species – I assume that it would not be a failure of just English speakers – not to have managed over many millennia to come up with a conventional way of conveying a piece of information that is so causally significant.

19.3.2 Descriptive Names and "the contingent **a priori**" *(Salmon, Schwartz)*

Nathan Salmon's "Naming and Non-Necessity" is a thorough and judicious examination of Kripke's

> startling claim … that certain sentences that are (semantically) true as a consequence of the way a name's reference was fixed by description are metaphysically contingent yet knowable *a priori*. (##1)

Names that have their reference fixed by a description (more precisely, by an attributively used description) are now usually called "descriptive names". They stand in contrast to the paradigm names discussed in Sect. 19.2.2 which have their reference fixed by a causal grounding. 'Jack the Ripper' is a favorite example of a descriptive name and it is one of Kripke's. It is the example I shall use. Kripke supposes that "the police in London use the name 'Jack the Ripper' to refer to the man, whoever he is, who committed all these murders, or most of them" (1980: 79). Salmon takes Kripke to be committed to the view that the following sentence is "(semantically) contingent yet *a priori* for the reference fixer (at the time of fixing)":

(3) If anyone committed such-and-such murders, then Jack the Ripper did. (##3)

Let us start with the contingency thesis. Salmon does not mince his words: "Kripke has persuaded the angels and all right-minded philosophers that [(3)] is indeed contingent" (##5). Well, as Salmon notes (##n. 6), Kripke has not persuaded me (even though named after a top angel). The thesis that (3) is contingent rests on the thesis that 'Jack the Ripper' is rigid. But why suppose that? To Salmon and other direct referentialists this rigidity thesis clearly seems intuitive, as intuitive as the rigidity thesis about paradigm names like 'Aristotle'. It doesn't seem so to me. I have noted that

> when we submit [descriptive names] to the standard tests for rigidity, they do not *clearly* pass, even if they pass at all. Consider, for example, how 'Jack the Ripper' fares on what Kripke describes as an "intuitive test" of rigidity (1980: 48–49, 62). The following should be false:
>
> Jack the Ripper might not have been Jack the Ripper.
>
> Yet it seems not to be so, at least not clearly so. Similarly,
>
> It might have been the case that Jack the Ripper was not a murderer,
>
> (with 'Jack the Ripper' having narrow scope) should be true, but it seems not to be. And suppose that Prince Alfred was Jack the Ripper. Is it really necessary that he was? If ever

> there was a thesis in the philosophy of language that needs more than intuitive support – and I think they all do (2012a, b) – the thesis that descriptive names are rigid is surely one. (2015b: 136–137)

Note that the intuition that a paradigm name like 'Aristotle' is rigid, an intuition that I do share, does have other support: it is rigid because its reference is fixed by a causal grounding in an actual object (2005a: 145). But there is no analogous support for the rigidity of 'Jack the Ripper'. Salmon quotes (##5) a passage in which Kripke imagines

> a hypothetical formal language in which a rigid designator '*a*' is introduced with the ceremony, "Let '*a*' (rigidly) denote the unique object that actually has property *F*, when talking about any situation, actual or counterfactual." (1980: 14)

Clearly descriptive names *could be* introduced into our actual language in this way (perhaps implicitly). But where's the evidence that they *are*? After all, we can imagine another hypothetical formal language in which a *non*-rigid designator is introduced like this: "Let '*a*' denote the unique object that has the property *F* in a situation, actual or counterfactual." (Note that this alone does not make such a designator *synonymous with* its introducing description. It might be open, as descriptive names like 'Jack the Ripper' *are* open (2015b: 136), to being borrowed by people who are ignorant and wrong about the referent.) I rather doubt that there is any determinate matter of fact about which of these hypotheticals fits our actual practice. And we surely don't have any evidence of which.

Turn now to the *a priori* thesis. Salmon has a traditional understanding of '*a priori* knowledge': "knowable with epistemic justification that is independent of experience" (##1); his example is justification by mathematical proof (##2). He carefully considers how reference fixing might yield a justification of the likes of (3).[55] He demonstrates that the justification rests on knowledge of semantic facts that is *a posteriori* (##7–8). He goes on: "We should probably conclude from the preceding considerations that Kripke means something different by his use of the term '*a priori*'" (##8). He comes up with a proposal:

> I submit that what Kripke has in mind by his use of the term '*a priori*' is a truth that is knowable independently of any experience beyond that on which knowledge of purely semantic (and/or purely pre-semantic) information about the language in question depends (even insofar as such knowledge is *a posteriori* in the traditional sense)…. I shall say that a truth is *quasi-a-priori* – for short, *qua-priori* – if it fits this broader notion. (##9).

Even an old Quinean like me who believes that there is no *a priori* knowledge at all (1998, 2011d) could accept that our knowledge of (3) is *qua-priori*.

Salmon has made a good point about the empirical nature of the semantic knowledge, arising from a stipulated meaning, that plays a role in the justification of the allegedly *a priori* (3). I have argued that we should go further along these lines: the semantic knowledge, supposedly arising from "conceptual analysis", that plays a role in the justification of allegedly *a priori* "analytic truths", is also empirical.

[55] His consideration is actually of another Kripke example, the one-meter stick. My remarks extend *mutatis mutandis* to a further example, Neptune, if not also to the one-meter stick.

Thus, our knowledge of “All bachelors are unmarried” is supposed to rest on our semantic knowledge that the content of the concept <bachelor> is the same as that of <adult unmarried male>. But that semantic knowledge, if knowledge it is, would come from empirical semantics (2011d: 25). Consider also this claim by Stephen Schwartz in “Against Rigidity for General Terms”: “People who are conversant with the terms ‘sloop’ and ‘sailboat’ know *a priori* that all sloops are sailboats” (##11). If those people have this knowledge about sloops on the basis of their semantic knowledge about ‘sloop’ and ‘sailboat’, then the knowledge is empirical.

In sum, the famous examples of “the contingent *a priori*” are not *a priori* and may not be contingent.

19.3.3 Rigidity in General Terms (Schwartz)

There is a well-known problem extending Kripke’s notion of rigidity from singular terms to general terms. Schwartz begins his paper with a nice description of this problem and of two proposed solutions to it. He calls one of these proposals “rigid expressivism”. It is the view that “rigid general terms designate or express the same kind or property in every possible world” (##2). Schwartz rejects this, as he has done before in “Kinds, General Terms, and Rigidity” (2002). I have also rejected it in “Rigid Application” (2005a: 140–143; also in 2009b, in response to Orlando 2009). I find his case against rigid expressivism largely persuasive and will not discuss it. Schwartz calls the other proposal “rigid essentialism”. He rejects this in general and my version in particular. Here is my version:

> a general term ‘*F*’ is a rigid applier iff it is such that if it applies to an object in any possible world, then it applies to that object in every possible world in which the object exists. Similarly for a mass term. (2005a: 146)

Now Schwartz and I both emphasize that a notion of rigidity must do some theoretical work. But we differ sharply on what that work is. Schwartz thinks that the job of rigidity is to distinguish “natural” kind terms from others. This led him to criticize an earlier presentation of my view (Devitt and Sterelny 1999: 85–86) on the ground that rigid application does not do that job (Schwartz 2002). He claimed, on the one hand, that some nominal kind terms like ‘television set’ are rigid appliers. He claimed, on the other hand, that some natural kind terms like ‘frog’ are not rigid appliers. My negative response to Schwartz was to argue that “even if these claims are right [I don’t think they are about ‘television set’ (2005a: 155–156)], they are not grounds for dissatisfaction” with my notion of rigidity because it is *not* the task of such a notion to distinguish natural from non-natural kind terms. I also argued, positively, that “the primary task is to distinguish kind terms that are not covered by a description theory from ones that are” (2005a: 154). In the present paper, Schwartz responds:

> Rigid essentialism offers no systematic or plausible distinction between natural kind terms and nominal kind terms nor is it useful in defeating descriptionism and thus loses its point. (##3).

So, despite my negative argument, Schwartz is still demanding that rigidity distinguish natural and nominal kind terms. And, despite my positive argument, he denies that rigid application counts against descriptionism. I shall briefly rehearse my arguments before turning to what Schwartz says in support of his position.

Positive Thesis On the basis of the theoretical work that Kripke did with rigidity in discussing proper names, I argued (2005a: 144–148) that "the primary work we should expect from a notion of rigidity for kind terms is featuring in lost rigidity arguments against description theories of meaning for some terms" (145). Lost rigidity arguments have the following form: the term in question is rigid; a description of the sort that the description theory alleges to be synonymous with the term is not rigid; so, the term is not synonymous with that description and the description theory is false. Finally, rigid application does indeed feature in lost rigidity arguments and so does the required work. So far as I can see, nothing Schwartz says actually counts against this thesis.

Suppose that the thesis is right. It raises two questions. (I) *What* terms are rigid and hence can feature in lost rigidity arguments? I shall discuss this soon. (II) What *other* theoretical tasks, if any, should rigidity perform? This brings us to the negative thesis.

Negative Thesis It is true that Kripke's favorite examples of rigid kind terms are (arguably) natural kind terms. Still, Kripke does not claim that all natural kind terms are rigid nor that all non-natural ones are non-rigid. Indeed, he thinks that "presumably, suitably elaborated…'hot', 'loud', 'red'" (1980: 134) are among the rigid ones; see also the discussion of yellowness (128n). So, in partial answer to question (II), I could see no basis for the idea that rigidity *should* mark out the class of natural kind terms (2005a: 145–146). Later, in a passage mostly quoted by Schwartz (##5), I've wondered why the idea that it should mark out this class

> has any *prima facie* appeal at all. First, … 'natural kind term' is vague. The problem is that it is far from clear what is for a *kind* to be natural and which ones count as natural…. Second, … it is hard to see how 'natural kind term' could come out as a theoretically significant description in semantics. Thus, 'plastic' is not likely to be classified as a natural kind term and yet it is surely semantically just like the paradigmatic natural kind term 'gold': the two terms seem equally nondescriptive …; and if there is any acceptable sense in which 'gold' is rigid then surely 'plastic' will be rigid in that sense too. Furthermore, the biological term 'predator' must be counted as natural and yet it seems descriptive and nonrigid. And what could be the principled basis for counting terms from the social sciences like 'unemployed' and 'nation' as not natural? Yet they are surely descriptive and nonrigid…. '[N]atural kind term' does not cut semantic nature at its joints; it does not describe a natural kind! (2009b: 245)

I now have a more crushing objection to the idea. How *could* rigidity have the *theoretically interesting* task of marking out the class of natural kind terms? It is trivial that a term is a natural kind term iff it refers to a natural kind. So, if we mark out the kinds, we mark out the terms; and vice versa. And we should proceed by marking out the kinds and thus the terms: we should "put metaphysics first", as I like to say (2010c); see Sect. 19.4.1. Indeed, it is somewhat preposterous to think that we should proceed in the other direction, marking out natural kinds by way of a semantic thesis about natural kind terms. And if there is any theoretical interest in marking out the terms at all, it is because there is one in marking out the kinds.[56] *Schwartz criticizes rigid application for failing to do something that nobody should have ever thought a notion of rigidity needed to do.*

What does Schwartz have to say in support of his position? He thinks that my "theory depends for its plausibility on a limited diet of examples." I "cherry pick" terms (##3). But this can't be an effective criticism of my positive thesis. According to that thesis, rigidity's task is to feature in lost rigidity arguments against description theories. So rigidity "can apply where it may without reflecting on its worth" (2005a: 146). Schwartz makes it immediately apparent what lies behind his criticism: he is insisting, despite my negative thesis, that rigidity *must* sort the natural from the nominal kind terms: 'frog' is natural but not a rigid applier[57]; 'refrigerator' is nominal but a rigid applier. He concludes: "Devitt's notion of rigid application cannot be used to systematically sort general terms into rigid-like natural kind terms versus non-rigid nominal kind terms" (##4). But, as I emphasized in my earlier papers, this sorting is *not* the theoretically interesting role for rigidity. So far, Schwartz is replaying his earlier objections (2002), to which I have already responded (154–159).

Let us say more about Schwartz's nice example of 'frog'. I presented his point that 'frog' is not a rigid applier thus:

> Consider a particular frog. It starts life as a tadpole and then turns into a frog. So 'frog' then applies to it. But in another possible world it dies young as a tadpole. So 'frog' never applies to it in that world. (2005a: 157).

I went on to argue for a view of 'frog' (157–159). I have since abandoned that view in responding to the criticisms of Ezequiel Zerbudis (2009) and have come up with a new one. This new view deploys a notion of *mature-rigidity*:

> A general term 'F' is a mature-rigid applier iff it is such that if it applies to an organism in any possible world, then it applies to that organism in every possible world in which the organism exists and develops to maturity.
> 'Frog' is a mature-rigid applier: anything that is a frog in some world will be a frog in any other possible world in which it exists provided it does not die as a tadpole. (2009b: 248).

[56] Schwarz makes the interesting suggestion that we follow Mill in taking natural kinds to be those "whose nature is inexhaustible" and nominal kinds those "which rest on one or only a few criteria" (##5). Suppose we do follow Mill, then we could *immediately*, without any appeal to semantic theory, distinguish natural from nominal kind terms!

[57] Schwartz also wonders whether it is possible for "mature eagles [to] turn into mature tigers" (##5). I agree with the biological consensus that this is not possible: it is essential for something to be a tiger that it has a certain history (2018). Were an eagle to turn into something tiger-ish that something would not be a tiger because it would not have the required history.

Schwartz is dismissive: "This strikes me as adding complication to complication in an *ad hoc* attempt to save a theory" (##4). I presume that Schwartz thinks this because he takes my claim that 'frog' is a mature-applier to be an "attempt to save the theory" that rigid application distinguishes natural kind terms. But that is not the purpose of the claim for that is not my theory. My negative thesis, in partial answer to question (II), is precisely that it is *not* the task of rigidity to distinguish natural kind terms. Schwartz has misunderstood the point of my discussion of 'frog'.

The discussion arises from my concern with question (I), with what terms are rigid and hence can feature in lost rigidity arguments. The point of the discussion was to *lessen the disappointment* of discovering, thanks to Schwartz, that rigid application did not yield a lost rigidity argument for 'frog' (2005a: 157). I was looking for *another* notion of rigidity that would yield such an argument for 'frog'. Mature-rigidity does the trick: it

> will serve to refute any description theory of 'frog' constructed in the usual way from descriptions of the readily observable properties of frogs. In the actual world those descriptions apply to frogs but in another possible world they might apply not to frogs but to other organisms altogether. The descriptions form a non-mature-rigid applier. (2009b: 248–249)

In sum, rigid application features in lost rigidity arguments for some terms, mature-rigid application, for others. Mature-rigidity is not introduced "to save a theory" but to add another one. And, to repeat, neither theory has anything to do with distinguishing natural kind terms.

I claim that the task of rigidity is to feature in lost rigidity arguments. I claim that rigid application and mature-rigid application both fulfil that task for some terms. Schwartz has not said anything that casts doubt on these claims. His criticism that these notions do not distinguish natural from nominal kind terms is misguided. It demands that rigidity do something that nobody should have ever thought it needed to do.

19.3.4 Narrow Meanings (Lycan, Horwich)

In his paper, "Devitt and the Case for Narrow Meaning", Bill Lycan generously describes my paper, "A Narrow Representational Theory of the Mind" (1989b), as "distinctive and valuable" (##1). Well, distinctive it may be, but I do wonder about its value. By the time of *Coming to Our Senses* (1996), I had come to realize, as Lycan notes (##16), that that paper rests on a serious conflation of the important distinction between narrow meanings as functions from contexts to wide meanings and as functional roles (255 n. 10). More on this distinction below.

Narrow meaning/content was all the rage in the 1980s. It was generally agreed that the meanings we ordinarily ascribe to explain behavior are truth-referential and hence "wide". Still, many thought that the meanings we *ought* to ascribe for that purpose should be "narrow". For some, cognitive psychology should explain the interaction of mental states with each other and the world by laws that advert only to formal or syntactic properties. For others, the semantics for psychology should be richer in some way but still narrow. I was one of those who took this revisionist line. I regret doing so.

My view on this issue in the 1989 paper is ably presented and criticized by Lycan (##1–15). Some features of that paper re-appear in the 1996 book, but placed in a framework that strictly observes the aforementioned distinction that was conflated in the paper. I still believe what I said about the issue in the book and that is what I shall discuss. That book, like the earlier paper, focuses on two arguments for revisionism, the argument from the computer analogy (1996: 265–272) and the argument from methodological solipsism (272–277). Lycan and I are in almost total agreement.

First, neither of these argument supports the view that psychology should ascribe only syntactic properties,[58] strictly understood, to mental states. Such properties may be adequate for the explanation of thought processes but they are not for the explanation of thought formation or behavior. The mind as a whole is not purely syntactic at any level even the implementational (1996: 277–284). In arguing this, *Coming*, like the 1989 paper, emphasizes three distinctions that tend to be overlooked in the debate: first, that between a token's intrinsic fairly brute-physical properties like a shape, which I call "formal" properties, and syntactic properties, which are extrinsic functional properties that a token has in virtue of its relations to other tokens in a linguistic system (258–265); second, that between processes that hold only between thoughts, and mental processes in general, which may involve not only thoughts but also sensory inputs and behavioral outputs (266–267); and, third, that between syntactic properties, which are constituted only by relations between linguistic tokens, and putative narrow meanings, which involve relations to nonlinguistic entities – for example, sensory inputs and behavioral outputs – as well (273–275).

The argument from methodological solipsism may seem to support the view that psychology should ascribe only narrow meanings. To assess this support we need to make that aforementioned distinction between two views of narrow meaning (285–286). According to one,[59] the narrow meaning of a sentence is a function taking an external context as argument to yield a wide meaning as value. So, on this view, narrow meanings partly determine truth conditions and reference; they are intentional wide meanings "minus a bit"; they are "proto-intentional". The belief that we need only these meanings to explain behavior is, therefore, only moderately revisionist. According to the other, more popular, view of narrow meaning,[60] that meaning is a functional role involving other sentences, proximal sensory inputs, and proximal behavioral outputs. These putative meanings are not truth-referential and differ greatly from the meanings that we currently ascribe. The belief that we need only these meanings to explain behavior is, therefore, highly revisionist.

[58] Cf. Stich 1983; but see also Stich 1991.

[59] To be found, e.g., in White 1982; Fodor 1987: 44–53.

[60] To be found, e.g., in Loar 1981 and 1982, McGinn 1982, Block 1986.

Coming finds some truth, though not enough, in the argument for the moderately revisionist view, but none at all in the argument for the highly revisionist one. My responses to these arguments reflected the influence of Tyler Burge (1986).

Start with the moderately revisionist view. What precisely are these proto-intentional narrow meanings? What constitutes their functions?

> Consider what must be the case for any token to have a certain wide meaning. Part of the answer is that the token must have some properties that it has solely in virtue of facts internal to the mind containing it. Those properties constitute its narrow meaning. So, if we had a *theory* of its wide meaning, we would know all that we needed to about its narrow meaning. The theory will explain the narrow meaning by explaining *the way in which* the mind must take a fact about the external context as the function's argument to yield a wide meaning as its value. (286).

Lycan is puzzled:

> I don't understand Devitt's allusion to "the mind" and what it "must take." It is the *theorist* whose task is to explain the way in which the external context combines with internal properties of a mental token to yield the token's wide meaning. (##16).

Let me clarify. Suppose that we are considering the narrow meaning of 'bachelor' and our theory tells us that the reference of 'bachelor' is determined descriptively. Then, the function that is 'bachelor''s narrow meaning

> will be partly constituted by some inferential properties that will constrain what it could refer to in a context; thus, perhaps 'bachelor' could refer to something in a context only if 'unmarried' refers to it in that context. (1996: 290)

Suppose so. Then our theory tells us that the way in which the function for 'bachelor', a part of the mind, takes a certain fact about the external context as its argument is partly via the function for 'unmarried', another part of the mind, taking a certain other fact about the external context as *its* argument. In contrast, suppose that we are considering the narrow meaning of 'Aristotle' and our theory tells us that the reference of 'Aristotle' is determined causally. Then a certain causal network is the external fact that the function for 'Aristotle', a part of the mind, takes as its argument, without any dependence on the functions for other terms. This "direct" way of selecting an argument is very different from the indirect way for 'bachelor'.

Given that a theory that explains wide meanings will explain these functions, such narrow meanings must be acceptable to someone who believes, as I do, in wide meanings. And these narrow meanings would indeed yield explanations of behavior. But there is a concern about this because many of these narrow meanings are likely to be "coarse grained" in that there is not much to them and "promiscuous" in that they can yield any of a vast range of wide meanings as values by changing the relevant external context as argument (287–292). Lycan offers a "rebuttal" but I think this arises from a misunderstanding:

> even if mental proper names are purely referential expressions, of course they do not *by themselves* contribute much to psychological explanation, nor would be expected to. They occur syntactically impacted in longer constructions, paradigmatically whole sentences. (##19)

And he points out that the different coarse-grained and promiscuous narrow meanings of pronouns can make significantly different contributions to psychological explanations. I agree with both these points, but they are not at odds with my proposal. To say that a meaning is coarse-grained is not to say that it has *no* grain; to say that it is promiscuous is not to say that it can take *any* value in a context. And, remember, we are contemplating that the coarse-grained and promiscuous terms may include not just proper names and pronouns but many mass terms, general terms, adjectives, and verbs. Just how many should be included "is an open question to be settled with the help of future theories of reference" (1996: 291).

I go on to argue (299–312) that these narrow meaning could serve psychological needs if the behaviors that needed to be explained were themselves only narrow and "proto-intentional". The objection to restricting psychology to explanations involving narrow meanings is then that *we do not need to explain only proto-intentional behavior*. For, that behavior is also likely to be coarse grained and promiscuous. The intentional behavior, giving water to Mary involves giving, water, and Mary. So too does the related proto-intentional behavior in *one* context, but that behavior might, in some other context, involve taking, kicking, or many other acts; *XYZ*, or many other stuffs; Twin Mary, or any other person. We certainly want to have the more discriminating intentional behavior explained *somewhere*. No good reason has been produced for thinking that it should not be explained *in psychology*.[61]

Coming is rather more critical of the more popular functional-role narrow meanings than Lycan conveys (##20–21). I argue (292–299) that these putative meanings are left almost entirely unexplained and mysterious. Even if they were not, we have been given no idea how such meanings *could* explain intentional behaviors (like giving water to Mary) and it seems very unlikely that they could. If they do not explain these behaviors, then revisionism requires that intentional behaviors be denied altogether; for if there are these behaviors and they are not explained by narrow meanings then it is not the case that psychology should ascribe only narrow meanings. We have been given no reason to deny intentional behaviors. The argument from methodological solipsism does nothing to solve these problems for functional-role meanings. This is very bad news for the highly revisionist doctrine that psychology should ascribe only these putative meanings. That doctrine has a heavy onus arising from the apparently striking success of our present practice of ascribing wide meanings to explain behavior. Why do these ascriptions seem so successful if they are not really? What reason have we for thinking that the ascriptions

[61] Horwich wonders how properties constituted by types of causal chains that are largely external to the mind *could* explain "any specific use of a word" and hence be meanings (##9). The use of words is, of course, just one sort of intentional behavior that these properties are supposed to explain. How *could* they do that? It's a long story (1996: 245–312; 2001: 482–490). A provocatively short version is: they can because *they do*; we have good reason to believe that we ascribe these properties to explain behavior; these explanations are successful; so what we ascribe are meanings.

that would be recommended by this revisionist doctrine would do any better? The doctrine has hardly begun to discharge its onus.[62]

19.3.5 The Use Theory (Horwich)

Horwich has presented his use theory of meaning (UTM) with great verve and clarity (1998, 2005). I have responded with two critical articles (2002, 2011a). Horwich's "Languages and Idiolects" is, in part, a response. I shall start with a brief recap.

Horwich's UTM takes a meaning to be

> an acceptance-property of the following form:– 'that such-and-such w-sentences are regularly accepted in such-and-such circumstances' is the idealized law governing w's use (by the relevant 'experts', given certain meanings attached to various other words). (2005: 28)

In assessing UTM and comparing it with truth-referentialism, I noted (2002: 114) that we are unlikely to make progress by considering "nonprimitive" words like 'bachelor' which the truth-referentialist might well think are covered by description theories. We need to focus on "primitive" words like proper names and natural kinds terms, the meanings of which are likely determined by direct relations to the world. (Sects. 19.3.1, 19.3.2, 19.3.3 and 19.3.4).

Considering these primitives, I argued that "the very same considerations that were devastating for many description theories are devastating for Horwich's use theory" (2011a: 205). Like description theories of reference, UTM suffers from a Kripkean "ignorance and error" problem (Sects. 19.2.1, 19.2.5). Horwich captured neatly the form of this problem for UTM: "members of a linguistic community typically mean exactly the same as one another by a given word, *even when their uses of it diverge*" (1998: 85–86; my emphasis). Horwich responded to this problem with the idea of *deference to experts*. (This process, with its epistemically demanding backward-looking intentions, should not be confused with the causal theory's reference borrowing discussed in Sect. 19.2.1.) I complained (2002: 118) that Horwich offered only some brief remarks to explain this deference. And I gave three reasons initially (2002: 118–119) and two more later (2011a: 205) for doubting that these remarks could be developed into a satisfactory theory of deference. In brief:

> (i) People will often not defer where they should. (ii) They will often try to defer but fail to identify an expert. (iii) They will often defer to a nonexpert. (iv) When the bearer of a name has been long-dead – for example, Aristotle – there will be no experts around to defer to. (v) Where there are surviving experts about a dead person, there seems to be no change that a

[62] An autobiographical note. Bill Lycan (##n. 9) mentions the often-hilarious episode of "Monty Python's Flying Circus" featuring the philosophy department at "the University of Woolloomooloo" (which led David Lewis to call his cat "Bruce"). There is of course no such university in Woolloomooloo, but there is a large pier. I sailed to England from that pier in 1947. Woolloomooloo used to be a poor area suitable for students. I lived there in 1965 when an undergraduate at the University of Sydney. Woolloomooloo is no longer poor and Russell Crowe lives on the pier.

> deferrer could be disposed to make to conform to the experts' basic acceptance properties. (2011a: 209)

Finally, I raised the "deeper problem" that "this appeal to deference seems to be incompatible with UTM" (2011a: 206).

> The problem for UTM arises from the following dilemma: either the members of a linguistic community typically share their meaning of a primitive word or they do not. Horwich's appeal to deference seemed to arise from his grasping the first horn of the dilemma. (2011a: 206).

But, I argued, UTM cannot combine deference with shared meanings. Further, I argued, "the trouble with the second horn is that it is false" (207): *meanings are typically shared*. I concluded: "These problems are not minor ones of details. The problems strike at the very core of UTM. At the very least, Horwich owes us an account of how UTM can deal with them" (209). I take Horwich's present paper to be attempting just that.

Horwich responds to my dilemma, in effect, *by grasping both horns*! The members of a linguistic community share one meaning of a word, the "communal" meaning (##2), but often not another, an "idiolectal" meaning (##3). Horwich thinks he can have it both ways! This is characteristically clever but, I think, provides no way out of the dilemma.

What about these idiolectal meanings? In talking about them in point (5), Horwich makes

> a distinction (albeit rather vague) between ordinary words and technical terms. In the former case, there's a basic rule for the word's use, which many of the speakers who often use it implicitly follow (and the majority follow at least approximately). (##3)

So with "ordinary words" idiolectal meanings are *more or less* shared. But with "technical terms" the story is very different:

> the members of some small subset of the population – acknowledged experts in the relevant area – nearly all implicitly follow a certain rule for its use, whereas the non-experts rarely follow it. (##3)

The "nonexperts" are, of course, those who feature in Kripkean ignorance and error arguments. These arguments show that many users of proper names, natural kind terms like 'elm', and perhaps lots of other terms are "nonexperts". So Horwich's "technical terms" must cover *all* these terms, indeed, all primitives and perhaps some nonprimitives like 'sloop' and 'arthritis' as well (2011a: 204–205). On his view, "experts" will have idiolectal meanings of these terms that will be very different from those of the "non-experts" and those of the "nonexperts" will differ from each other.

What about the communal meanings? Talking about them in point (3), Horwich says

> Each word's sound-type … has a constant *communal* meaning. That type-meaning – which, together with the *context* of an utterance containing the word, determines the word's contribution to *what is said* – does not vary from one speaker to another. (##2–3)

This is vital to the plausibility of UTM, of course, because to suppose that a group of organisms has a language at all is to suppose that they share meanings. But how is this sharing possible given the deficiencies of the nonexperts? In point (6) Horwich says that

> such deficiencies – no matter how great – don't count against attributions of the communal meaning. Individuals are credited with that communal meaning simply by virtue of their membership in the community. (##4)

All this raises two crucial questions which I will get to in a moment. But first I should briefly state my own position on the relation of idiolects to communal languages (2006a: 178–184; Forthcoming-b: ch. 5). For, Horwich rightly claims that my idiolects "are crucially different" from the idiolects that he finds so theoretically important (##12 n. 10).

Idiolects The meaning of an expression in a person's idiolect is *constituted* by a linguistic rule. I take that rule to be the person's disposition to associate the expression with that meaning in the production and comprehension of language: she is disposed to use that expression with that speaker meaning; and she is disposed to assign that meaning to the use of that expression by others who she takes to share her idiolect. Occasionally a person will not do what she is normally disposed to do; she will deliberately assign another meaning to an expression, as in a metaphor or pragmatic modulation; or she will make a performance error. In these cases, an expression will have a speaker or audience meaning that is different from its literal meaning in the person's idiolect.

What *causes* the expression to have its meaning in a person's idiolect? (i) If Chomsky is right, as he probably is, some syntactic rules are *innate*. So part of the meaning of an expression may be innate. (ii) But a large part of its meaning is typically the result of the person's participation in a *convention*; so, it is a conventional meaning. (iii) Finally, some of a person's idiolect may be her own work and so a bit idiosyncratic: the literal meaning an expression has for her may not be a meaning it has according to any linguistic convention; Mrs. Malaprop is a famous example.

Languages An idiolect is shared in a community, and hence is the language of that community, to the extent that the members of the community are disposed to associate the expressions of the idiolect with the appropriate meanings. This sharing could, in principle, come about "by chance" as the result of each member's own work. But, of course, it does not. It probably partly comes about because of some innate syntax. It is largely the result of the community participating in the same linguistic conventions. So, conventions are the typical *cause* of communal meanings (but, if anything close to this story is right, it is a mistake to suppose that conventions *constitute* meanings).

What is it for the communal meaning of an expression to arise from a convention in a community and hence be a conventional meaning? It is hard to say, and I must be brief anyway. The central idea is that each member of that community is disposed to associate the expression with that meaning *because other members are similarly* *disposed*; there is a certain sort of causal dependency of the disposition of each member of the community on the dispositions of others. The norm is for speakers in a community to share a linguistic meaning because they stand in the required causal relation. As a result, the idiolectal meaning of almost all expressions for almost all speakers will be the conventional meaning of those expressions in the speakers' community. Mrs. Malaprop's divergence from the norm is exceptional.

Take any word. On this picture, a person usually participates in the conventions for that word in her community, but she may not. If she does participate, then her idiolectal meaning will be the same as the communal meaning. If she does not participate, it will be different; the word out of her mouth will literally mean something other than its meaning out of the mouths of her fellows who participate in the convention. But the word out of her mouth cannot literally have *both* her idiolectal meaning *and* the different communal meaning. Having both is simply impossible. Yet, on Horwich's view of idiolectal meanings, this is not just possible but so common as to be almost the norm. I can see no theoretical justification for positing idiolectal meanings that are so unconnected to communal meanings; I do indeed think, as he supposes, that they are "completely beyond the pale" (##12 n. 10).

Horwich calls the acceptance property of an individual's word her "idiolectal meaning". That property is real enough but for it to be *properly* called a "meaning" at all, as Horwich and I agree, it has to play a certain causal role (2011a: 197–198). And any property that plays that role has to be related to what is properly called a "communal meaning" in something like the way I have described. Horwich's acceptance property does not meet that requirement.

It is time for the first crucial question. *In virtue of what* is a *certain* acceptance property the communal meaning of a word? What makes it the case that in a community of idiolects varying from experts to non-experts and among non-experts, *that property* is the meaning of the word in the community? Horwich has a breezy answer in his point (8):

> The upshot is that the meaning of a word in a communal language is grounded in the meanings it has within the various idiolects of individual members of the community – in the way that was indicated in points (5) and (6). (##4).

Horwich says later: "As for the name's *communal* meaning, this might well be grounded in the basic acceptance rule implicitly followed by the relevant experts" (##6–7). The idea that comes through from this is clear enough: with "technical terms", which need to be all terms for which ignorance and error problems arise (including proper names), the communal meaning is the experts' idiolectal meaning. And the "grounding" of communal meaning involves deference. But how does this work? We need an explanation of this grounding.

Set that aside for a moment and consider a second related question, just as crucial as the first. It will be remembered that Horwich claims that, despite their deficiencies, nonexperts are credited with the communal meaning "simply by virtue of their membership in the community". But how could that be right? For them to be members of the community, *in the respect that matters*, they have to participate in the linguistic conventions of that community. It is not enough that they just hang out together: they have to be *talking to each other in the same language*. And the non-experts can't be doing that because they have idiolectal meanings that are quite different from the communal conventional meaning. Insofar as they are members of a linguistic community, hence participating in the conventions of the language, the communal meanings *are* their idiolectal meanings. What could there be about these nonexperts that makes a communal meaning *theirs*, given that they have various idiolectal meanings different from the community meaning and so are not participating in the communal conventions? How could their words be *correctly* credited with both meanings?

Something like Horwich's old idea of deference is again supposed to provide the answer. I have three responses to this. (1) I still have my old complaint that we have nowhere near enough explanation. He now thinks it was a mistake to suggest that "in order for a non-expert to use a technical term with its communal meaning she must *defer* to what the experts say with the help of that term." So he weakens his deference requirement: "The only role now given to this notion is in roughly explaining the notion of the 'acknowledged experts' as 'those to whose opinions there is a *tendency* to defer'" (##7–8 n. 6). But he says no more. (2) *Prima facie*, that weakening does not escape the five objections summarized above. (3) For the reasons just given, there is no legitimate distinction between idiolectal and communal meanings that can solve his "deeper problem": his appeal to deference seems to be incompatible with UTM; UTM cannot account for the sharing of meanings in a community.

I concluded "Deference and the Use Theory":

> In my view, the best hope for UTM would be to make it part of a hybrid account: combining UTM for the experts with the causal theory of deference [reference borrowing]. So the meanings of nonexperts will be determined by the meanings of the experts by means of deference as explained by the causal theory. At bottom meaning would be explained by use but otherwise not. It would be interesting to explore the problems for this hybrid. (2011a: 209).

Just as there is reference fixing and reference borrowing (Sects. 19.2.1, 19.2.2, 19.2.3, 19.2.4 and 19.2.5), there is meaning fixing and meaning borrowing. My hybrid proposal combines a use theory of meaning fixing with a causal theory of meaning borrowing. (My own view combines a *referential* theory of meaning fixing with that causal theory of borrowing.) Horwich has not, of course, embraced this proposal but some of his remarks suggest that he is not far away from it. Thus he speaks approvingly of what he aptly calls Kripke's "'contagion' picture of communal-reference inheritance" and goes on:

> But could it be that S's merely *hearing* the name, 'N', from a fellow community member will, by itself, guarantee that, when she proceeds to use it, 'N' will have the very communal meaning and referent with which her source deployed the word? I'm inclined to say 'yes'. (##7).

This sounds awfully like the causal theory of reference/meaning borrowing (Sect. 19.2.1). If he'd only drop his talk of "tendency to defer", and drop his idiolectal meanings that are beyond the pale, he would have the hybrid. I don't think the hybrid is right but it has much better prospects than UTM.

19.4 Methodology

19.4.1 Putting Metaphysics First (Rey)

I have already discussed the parts of Rey's "Explanation First! The Priority of Scientific Over 'Commonsense' Metaphysics" that concern linguistics (Sect. 19.1.1). The main concern of his paper is methodological. Rey has two criticisms of what he takes to be my methodology. The first of these is fairly harmless, the second is not. Both criticisms are baseless. I shall discuss the first in this section, the second in the next.

I am fond of the maxim, "Putting Metaphysics First". Rey thinks that it should be replaced by the maxim of putting *explanation* first because that maxim is more "fundamental".

Here is what I say about my maxim in a book named after it:

> We should approach epistemology and semantics from a metaphysical perspective rather than vice versa. We should do this because we know much more about the way the world is than we do about how we know about, or refer to, that world. The epistemological turn in modern philosophy, and the linguistic turn in contemporary philosophy, were something of disasters in my view. My view here reflects, of course, my epistemological naturalism. The metaphysics I want to put first is a naturalized one. (2010c: 2).

The book is an extended argument for this point of view.

Rey has this to say in favor of his maxim:

> I'd like to suggest replacing Devitt's maxim with what seems to me a more fundamental one: *Explanation first!* I think it's our reliable sense of explanation that underlies his and many other people's considered judgment that, at least for the time being, the "metaphysics" of much of the natural sciences ought to enjoy a priority over most all other claims, metaphysical, semantic, epistemic or otherwise.... I don't expect Devitt to seriously disagree with any of this.... [H]e has in mind the metaphysics of the explanations provided by good empirical scientific theories. (##2).

Indeed I do not disagree (apart from his replacement suggestion). All my work reflects an enthusiasm for using inference to the best explanation, or "abduction", to find out about the world; thus, in discussing meaning I say, "As always, we seek the best explanation" (1981a: 115; see also, 1996: 48–86); my argument for scientific realism is an abduction (1984: 104–106; 1991: 108–110); my position on linguistics (2006a), which so appalls Rey – see Sects. 19.1.1 and 19.1.2 – is driven by explanatory concerns. But Rey's maxim urging explanation as fundamental in epistemology could not be a *replacement* for my maxim because the two maxims have different concerns. My maxim is aimed at reversing practices of deriving a metaphysics from assumptions in epistemology or semantics, practices that have dominated the last three centuries. Rey's maxim offers no guidance on those practices. And mine offers no guidance on what is fundamental in epistemology. Rey is making a false contrast. This is surely obvious and so Rey's claim that his maxim is "a substantive alternative" to mine (##19) is very odd.

Rey's second criticism of my methodological is much more serious, so serious that Rey wonders whether it is "a serious indictment of Devitt's work as a whole" (##23)!

19.4.2 *"Moorean Commonsense" (Rey)*

Rey and I have aired disagreements on a number of substantive issues in many publications. In his present relentlessly unsympathetic paper, Rey seeks to discredit my side on these issues by attributing it to bad methodology: I am alleged to be steeped in the vice of "Moorean commonsense". He, in contrast, is full of the virtue of science:

> In Devitt, I tentatively want to suggest it is an uncritical, essentially Moorean "commonsense" understanding of realism and of Quine's "holistic" epistemology, which, I argue, badly biases his views of many topics he discusses, specifically (what I'll discuss here) secondary properties, linguistics, and the possibility of *a priori* knowledge. But all of this will be on behalf of stressing what seems to me to be the priority of scientific explanation over commonsense (##3).

What starts as a tentative suggestion soon becomes the dominant thesis of Rey's paper.

Let's start with Rey's criticism of my inclusion, in *Realism and Truth* (1984, 1991), of "commonsense" physical entities in my definition of realism: "Devitt … seems to include in the metaphysics he's putting first, 'most' commonsense claims, which, I argue, have no place in serious explanation" (##22). So, why do I include these claims about the existence of commonsense entities? Because I aim to reject a much broader range of antirealisms about "the external world" than antirealisms about science. This range includes idealisms, so prominent in the history of philosophy, that deny the independent existence of a world consisting not only of scientific entities but also of more humdrum ones including, for example, chairs and hammers (1991: 246–249). I describe one such doctrine as follows:

> *Constructivism*. The only independent reality is beyond the reach of our knowledge and language. A known world is partly constructed by the imposition of concepts. These concepts differ from (linguistic, social, scientific, etc.) group to group and hence the worlds of groups differ. Each such world exists only relative to an imposition of concepts. (1991: 235).

This sort of neo-Kantian relativism is common among intellectuals. Indeed, it has some claim to be the metaphysics of the twentieth century, exemplified by theorists ranging from Nelson Goodman to French feminists (1991: 235–236). And it has had consequences in social and political life, bad ones in my view. Perhaps Rey thinks that such antirealisms are not worth refuting. Well, I disagree. It's dirty work but someone has to do it. (See Sect. 19.5.1 for more on the definition of "realism".)

That criticism concerns what topics are worth discussing. The next is much more important because it concerns the methodology for discussing any topic:

> Devitt's only real motivation for insisting on color realism is simply a wistful attachment to commonsense…. My point here is not to go to the wall upon denying the reality of all these things; I only want to insist upon the methodological point that it should be *scientific* explanation, not commonsense, that should be the arbiter of the issue. (##6).

> So what might be the argument for Devitt's [rejection of the psychological conception of linguistics]? It's hard to resist the impression that it's yet another instance of his commitment to the Moorean commonsense that seemed to be driving his views about color. (##8)

> despite his claimed commitment to a "naturalized epistemology," Devitt, like his mentor, Quine, seems to display a peculiar aloofness to the sciences of the domains he discusses…. It often seems as though he and Quine remain(ed) "outside" of the sciences they otherwise esteem, precisely as philosophers traditionally have, listening in from afar, overhearing snatches of their claims, arguing with them from a commonsensical point of view, but not engaging directly with the enterprises themselves. (##22–23)

This is stuff and nonsense, a picture of my methodology with little relation to reality.

Rey's charge is that my methodology is one of "commitment to Moorean commonsense". In fact, my attitude to common sense, or "folk theory" (as I mostly call it), is very different. My attitude is a consequence of my Quinean epistemological naturalism (1998, 2011c) and is briefly as follows. Folk theory can be a helpful place to start in the absence of science. We then look to science to discover whether folk theory, so far as it goes, is right. And we look to science to go further, much further. Some past folk theories have turned out to be spectacularly wrong. Still, given that conservatism is among the theoretical virtues (Quine and Ullian 1970: 43–53), being in accord with common sense is an advantage for a theory, though, of course, very far from a decisive one. My most explicit and detailed presentation of this attitude to folk theory and common sense is probably in *Language and Reality* (Devitt and Sterelny 1999: 286–287). The attitude is at least implicit in my many discussions of intuitions (e.g. 1996, 2006c, 2012a, 2014a)[63] and exemplified in all my work, from its beginning on reference (1981a), through work on realism (1984, 1991), meaning (1996), linguistics (2006a), and biological essentialism (2008d), to recent work on experimental semantics (e.g. 2011b, 2012b; Devitt and Porot 2018) and to Sect. 19.2.3 of these very responses.

In the face of all this obvious and apparently overwhelming evidence that I am very far from a devotee of "Moorean commonsense", what does Rey offer in support of his claim to the contrary? Well, there are some uncharitable speculations about my motivations ("wistful attachment to commonsense", "hard to resist the impression … of his commitment to the Moorean commonsense", etc.), but what we need is evidence that I engage in *arguments resting on "Moorean commonsense"*. Rey produces none. True, I do use the expressions 'Moorean' and 'common sense'/'commonsense' *separately* quite a few times in arguments in *Putting Metaphysics First* (2010c). But these uses do not provide the needed evidence.

[63] There is a small irony in Rey's criticism. I have, in effect, criticized as unscientific a linguistic methodology (2006a, 2006c, 2010a, b, 2014a, 2020) that Rey (2013, Forthcoming-a, Forthcoming-b) defends: the practice of relying on speaker's meta-linguistic intuitions as evidence.

Let us start with my use of ‘Moorean’. This term is used *only* to describe a certain strategy in argument. Here is one example. I start my case for putting metaphysics first, hence for *Realism*, by arguing that “*Realism* is much more firmly based than the epistemological theses … that are thought to undermine it” (2010c: 62). (This is the only example of my alleged Mooreanism that Rey quotes (##4).) I began calling this “a Moorean response” some time back when Steven Hales pointed out that it exemplifies Moore’s famous strategy of responding to a *modus ponens* with its corresponding *modus tollens*. Importantly, *this use of ‘Moorean’ involves no commitment to common sense*, as indeed the following note about the strategy makes clear:

> Note that the point is not that *Realism* is indubitable, to be held “come what may” in experience: that would be contrary to naturalism. The point is that, prima facie, there is a much stronger case for *Realism* than for the [epistemological] speculations. (2010c: 109 n. 19).

I finish my case for putting metaphysics first, hence for Realism, by appealing to naturalism.

Turn now to my use of ‘common sense’. My *only* use of it in an argument is in the one for *Realism* that Rey quotes (##4). *In the course of* my *initial* argument for *Realism* I claim that this *Realism* about ordinary objects is “the very core of common sense” (2010c: 62). And, I might have added, being in accord with common sense is a plus for *Realism*, given the virtue of conservatism. However, even this initial argument does not *rest* its case for Realism on common sense. And this initial argument is treated as far from conclusive: “the Moorean response is not of course sufficient” (2010c: 63). Indeed, how could it be? “Realism might be wrong: it is an overarching empirical hypothesis in science” (1991: 20). So, I follow up with a naturalistic argument for it (2010c: 63–66; see also 1991: 73–82).

In sum, my separate uses of ‘Moorean’ and ‘common sense’ provide no basis at all for attributing to me the methodology that Rey disparages with such relish. In particular, I have always embraced “the priority of scientific explanation over commonsense” (##3). Contrary to Rey’s charge, there is not a single argument in my work that rests on a “commitment to Moorean commonsense”.

Finally, the charge that I “display a peculiar aloofness” to the sciences I discuss (##22) is gratuitous.

I conclude with some brief comments on the three substantive disagreements to which Rey applies his mistaken thesis about my methodology.

Linguistics I have already responded to Rey’s discussion of linguistics (Sect. 19.1.1). I disagree with him not because I practice the methodology that Rey disparages but because I think that he is wrong about what the science shows.

Color Rey is very dismissive of my (somewhat tentative) neo-Lockean view of colors. The issue has never been important to me and my discussions of it are old and very brief (1984: 69–71; 1991: 249–251).[64] Perhaps if I revisited the issue more thoroughly I would revise my view in the face of contemporary vision science. I obviously agree with Rey that "it should be *scientific* explanation, not common-sense that should be the arbiter of the question" (##6).

The a Priori Rey's disparagement climaxes: "patently preposterous.… [A] mere mantra, a check written on a non-existent bank" (##20). Our mentor Quine is also a target. I refer any reader interested in my criticisms of the a priori to my 1998 and 2011d.

19.4.3 Intuitions (Martí, Sterelny, Jackson)

How do philosophers of language tell which theory of language is right? Edouard Machery, Ron Mallon, Shaun Nichols, and Stephen Stich ("MMNS") noted (2004) that theories of reference are tested by consulting referential *intuitions* and that those intuitions are nearly always those of the philosophers themselves "in their armchairs". Indeed, the consensus is that consulting intuitions is the method for the philosophy of language in general.[65] Consider this statement, for example:

> Our intuitive judgments about what *A* meant, said, and implied, and judgments about whether what *A* said was true or false in specified situations constitute the primary data for a theory of interpretation, the data it is the theory's business to explain. (Neale 2004: 79).[66]

MMNS responded critically to this consensus methodology by testing the intuitions of the folk, in particular those of undergraduates in Rutgers and Hong Kong, about cases like Kripke's Gödel and Jonah ones. Whereas philosophers of language, we can presume, almost all share Kripke's antidescriptivist intuitions about these cases, the experiments revealed considerable variation in the intuitions of the undergraduates; indeed, the intuitions of Hong Kong undergraduates leaned toward descriptivism in the Gödel case. MMNS presented this result as casting doubt on philosophers' intuition-based methodology for theorizing about reference and hence, of course, on the resulting theories of reference.

Genoveva Martí (2009, 2012) and I (2011b, 2012b, c) criticized MMNS for testing the wrong thing. Experimentalists should not be testing theories against *anyone's* referential intuitions but rather testing them against the reality that these

[64] Rey misrepresents an essay of mine, "Global Response Dependency and Worldmaking" (2010c: 122–136), as an "extended discussion of color" (##6, n. 10). In fact, the essay is not a discussion of color but of the consequences for realism of the view that all properties are "response-dependent". One color, redness, does feature but only as a "popular and plausible" *example* of a response-dependent property (122). I do not argue that it is a response-dependent property.

[65] And for philosophy in general, of course. Still some disagree with this consensus: see Deutsch (2009) and Cappelen (2012); for a response, see Devitt (2015a).

[66] In his latest work, Neale urges a very different and, in my view, much more appropriate view of the role of intuitions (2016: 231, 234).

intuitions are about. We do not rest our biological theories on evidence from the intuitions of biologists about living things, let alone from the intuitions of the folk. We do not rest our economic theories on evidence from the intuitions of economists about money and the like, let alone from the intuitions of the folk. No more should we rest our semantic theories on evidence from the intuitions of philosophers about reference and the like, let alone from the intuitions of the folk. *The primary evidence for a theory about a certain reality comes not from such intuitions about the reality but from more direct examinations of that reality*. The reality of reference relations is to be found in linguistic usage. So theories of reference need to be confronted with direct evidence from usage.

Martí and I are very much on the same side in our view of experimental semantics. Positively, we think that theories of reference should be tested against usage. Negatively, we are critical of the practice of testing against referential intuitions. As we are both fond of saying, the right response to armchair philosophy is not to pull up more armchairs for the folk.[67]

I shall explore this, and the crucial distinction between testing against referential intuitions and usage, in the next section. But first I want to say more about referential intuitions.

Talking about the consensus methodology in "Experimental Semantics, Descriptivism and Anti-descriptivism. Should We Endorse Referential Pluralism?", Martí remarks:

> The input the semanticist reflects on is actual usage: the 'Feynman' and 'Columbus' cases that Kripke uses to illustrate the ignorance and error arguments, are real cases, things that, according to Kripke, one hears in the marketplace. (##6).

What does Martí have in mind as reflecting on usage? If she is to be right about the actual philosophical methodology exemplified by Kripke, she should have in mind immediate reflection on an observed utterance that yields *a referential intuition*, a meta-linguistic judgment that that expression out of the mouth of that person refers to a certain object. But I fear that this is probably not what she has in mind, as we shall see in the next section.

Consider what Kripke uses as evidence in his discussion of 'Feynman'. He thinks that

> the man in the street [who could not differentiate Feynman from Gell-Mann] may still use the name 'Feynman'. When asked he will say: well, he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name 'Feynman' as a name of Feynman. (1980: 81)

The last sentence expresses a referential intuition. And it is surely an intuition about remembered actual utterances, as Martí suggests: Kripke has observed ignorant people using 'Feynman' and judges that they successfully referred to Feynman. Kripke's discussion of the reference of 'Columbus' similarly rests on such an intuition. He points out that laypeople's likely "misconceptions" about Columbus

[67] As Martí points out (##6), neither of us know the source of this witticism.

should mean, according to the description theory, that their uses of 'Columbus' really refer to somebody other than Columbus. Kripke's intuitive response is: "But they don't" (1980: 85).[68] This practice of testing a theory against intuitions is very different from one of testing it against usage, as we shall see.

What lies behind the standard practice of relying on intuitions? The answer suggested by the literature is that competent speakers of a language have some sort of privileged access to the referential facts about their language. I argue that this answer is mistaken (Devitt 1996: 48–54, 72–85; 2006a: 95–121; 2006c, 2012a). Intuitions about language, like intuitions in general,

> are empirical theory-laden central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning. (2006a: 103; 2006c: 491)

A speaker's competence in a language does, of course, give her ready access to the *data* of that language, the data that the intuitions *are about*, but it does not give her privileged access to the *truth* about the data.

My talk of "theory-laden" here can mislead. In "Michael Devitt, Cultural Evolution and the Division of Linguistic Labour", Sterelny doubts whether it is helpful to characterize people's intuitions about a domain as "their theory of that domain" (##5). Perhaps not. Still we might well think of them, as Martí does, as "theoretical; … the first step that the theorist engages in" (##6). And we should see them as mostly *the product of experiences* of the linguistic world. They are like "observation" judgments; indeed, some of them *are* observation judgments. As such, they are "theory-laden"' in just the way that we commonly think observation judgments are. That is, we would not make any of these judgments if we did not hold certain beliefs or theories, some involving the concepts deployed in the judgments. And we would not make the judgments if we did not have certain predispositions, some innate but many acquired in training, to respond selectively to experiences. So 'theory' in 'theory-laden' has to be construed *very* broadly to cover not just theories proper but also these dispositions that are part of background expertise.[69]

It is *not* a methodological consequence of my view that referential intuitions should have no evidential role in theorizing about reference. It *is* a consequence, however, that they should have that role only to the extent that they are likely to be reliable, only to the extent that they are *reliable indicators*. Sterelny compares referential intuitions unfavorably with knapping intuitions and doubts that the referential ones are reliable: "the reasons we have to trust intuitions seem largely absent when we consider agent's judgements about reference" (##6). In "Language from a Naturalistic Perspective", Jackson takes intuitions to be "beliefs coming from the exercising of recognitional capacities" (##3). This is an apt way to take them. He

[68] For more evidence of Kripke's reliance on referential intuitions, see Devitt 2015a.

[69] For more on "theory-laden" see Devitt 2012b: 19; 2015c: 37–38.

rightly sees himself as more sympathetic to the role of intuitions than I am.[70] For, the recognitional view still leaves a reliability problem. Is a person claiming to recognize an utterance as a case of referring to *x* reliable about such matters? That needs to be assessed empirically using independent evidence about reference relations obtained from linguistic usage. But, as noted, we need that independent evidence anyway.

Are the referential intuitions of philosophers likely to be more reliable than those of the folk? Sterelny is dubious again: "I doubt whether philosopher's intuitions have much more weight" (##6). Now, it is a methodological consequence of my view of intuitions that, insofar as we rely on referential intuitions rather than usage as evidence, we should prefer those of philosophers to those of folk because philosophers have the better background beliefs and training: they are more expert.[71] This is not to say that the folk may not be expert *enough* in judging reference in a particular situation. Whether or not the folk are is an empirical question which the theory does not answer. The theory tells us that "the more expert a person is in an area, … the wider her range of reliable intuitions in the area" (Devitt 2010b: 860) but this provides no guidance as to the level of expertise required for a person to be reliable about any particular sort of fact. Still we can say that, from the perspective of the theory, it would be no surprise if the folk, though perhaps reliable enough in judging reference in humdrum actual situations, were rather poor at doing so in fanciful hypothetical cases like Gödel (Devitt 2011b: 421, 426; 2012b: 25–26).

Why does Sterelny doubt that we should prefer philosophers' intuitions?

> It is true that philosophers are more practiced: they spend a lot of time and thought thinking about problematic hypothetical cases. But practice only leads to greater skill when coupled to feedback about success and failure …. Philosophers' practice in thinking about referential examples is not coupled to any correcting feedback loop (##6)

This concern about lack of reliable feedback is empirically well-based (Weinberg et al. 2010: 340–341). Still, as I have pointed out, "philosophers of language are confronted informally by language use that does provide feedback" (2012b, 25; see also p. 29).

It is time to turn to the experimental evidence and testing against usage.

[70] I take it that his greater sympathy stems from his view that the intuitions in question are pieces of "conceptual analysis" and are a priori (1998).

[71] This line of thought yielded an example of what has become known as 'the Expertise Defense' against the findings of MMNS. The Expertise Defense has led to a lively exchange of opinion: Weinberg et al. 2010; Machery and Stich 2012; Machery et al. 2013; Machery 2012a; Devitt 2012b; Machery 2012b; Devitt 2012c. See also the following exchange arising out of my analogous claim that we should prefer the grammatical intuitions of linguists over those of the folk: Devitt 2006a: 108–111; 2006c: 497–500; Culbertson and Gross 2009; Devitt 2010b; Gross and Culbertson 2011.

19.4.4 Experimental Semantics (Martí, Sterelny)

First, what is the linguistic usage in question? It consists of the largely subconscious processes of producing and understanding linguistic expressions; of moving from a thought to its expression and from an expression to its interpretation.[72] These processes are very speedy and might well be called "intuit*ive*". But, even if they express judgments – and many may express other mental states – the expressions *that we should primarily count as evidence* are not expressions of *referential* judgments, hence not expressions of *referential* intuit*ions*; rather they are expressions of judgments about *the nonsemantic world*.

Next, what is it to test a linguistic theory *T* against usage? Theorists reason as follows. *T* predicts that in condition *C* competent speakers will utter *E*. Then if speakers do, that confirms *T*; if speakers do not, that disconfirms *T*. Similarly, if *T* predicts that speakers will *not* utter *E* in *C* and they do, that disconfirms *T*; if they do not, that confirms *T*. This is testing *T* against usage.

How can we test theories of reference against usage? One way is to look to the corpus, the linguistic sounds and inscriptions that competent speakers produce as they go about their ordinary business without prompting from experimenters. Ironically, many of the vignettes used by experimentalists, from MMNS onward, to test theories of reference, vignettes that are all parts of the corpus, *themselves* provide evidence against "classical" description theories, independent of anything that the experiments using these vignettes show about reference. For, they include expressions that the description theory predicts a competent speaker would not be disposed to utter (Devitt 2012b: 27–28, Devitt 2015c: 47–49, Devitt and Porot 2018: 1562). Still the theorist cannot count on such windfalls; getting evidence about reference from the corpus is usually going to be very hard. Fortunately, there is another way to get evidence from usage: we can apply the technique of elicited production, taken from linguistics (Thornton 1995: 140). We describe situations and prompt subjects to say something that the theory predicts they will or will not say. Then we reason in the testing-usage way, as described in the last paragraph.[73]

Now a theorist who reasons in this way is "reflecting on usage", in some sense, but this reflection is very different from the sort, described in the last section, that yields referential intuitions. For, taking observed usage as evidence *by the sort of reasoning just described* is very different from the practice of taking that usage as evidence *by making immediate judgments about its reference*. And this difference is important in assessing Martí's claim, quoted in the last section, about what semanticists do. For, though philosophers probably do *some* informal testing of theories of reference against usage in the way described, as noted at the end of the last section, what they *primarily* do is test theories against referential intuitions (2015c: 39–43).

[72] So I include understanding under "usage".

[73] Linguists also get evidence from usage by testing reaction times, eye tracking, and electromagnetic brain potentials.

Our discussion of Kripke illustrates the latter. In passages like the following, Martí seems to have a different take on what Kripke does:

> There is no doubt that people are referring to Feynman when they use 'Feynman' even if they do not attach a uniquely identifying description to the name. And there is no question that people are referring to Columbus, and not to some Viking that in the 11th century set foot in the New World, when they use 'Columbus'. These, and other similar real cases of usage that Kripke mentions, provide the data that supports the conclusion against descriptivism. (##13).

What Martí finds indubitable here are referential intuitions, and Kripke is testing the description theory against those. Yet Martí seems to take Kripke to be testing the theory against usage. (I hope this appearance is misleading.) For Kripke to be actually testing the theory against usage, say an observed use of 'Feynman', he would have to reason like this: the theory predicts that a competent user of 'Feynman' would not utter the name in circumstances *C*. X is a competent speaker and did utter 'Feynman' in *C*. So this count against the theory. This is manifestly not what Kripke is doing.[74]

Nicolas Porot and I have recently conducted experiments using elicited production to test classical description theories of proper names against usage (Devitt and Porot 2018). Our experiments were on Gödel and Jonah cases. The results were decisively antidescriptivist. So too were those of Domaneschi et al. (2017) who also used elicited production. We now have powerful evidence from usage in support of Kripke's antidescriptivism about proper names.[75]

I noted Sterelny's doubts about the reliability of the folk's referential intuitions. Experiments had given reasons for such doubts even before these recent tests of usage. We summed these reasons up as follows:

> First, these intuitions have proved quite susceptible to wording effects. Although the results of MMNS have been replicated several times, it has been found that small changes in wording yield somewhat different results.[76] Second, there is worrying evidence that the experimental task may be beyond many participants. The MMNS prompt asks participants to say who "John", a character in their vignette, "is talking about" when he "uses the name 'Gödel'". In one experiment (Sytsma and Livengood 2011, pp. 326–7), participants who had answered this question were then asked how they had understood the question: Is it about who John *thinks* he is talking about or about who John is *actually* talking about? Remarkably, 44 out of 73 chose the former, providing clear evidence that they had

[74] For more on the distinction between testing against referential intuitions and against usage, see Devitt 2015c: 39–45.

[75] Sterelny mentions an earlier elicited production experiment that Kate Devitt, Wesley Buckwalter and I used to test usage. Sterelny thinks that "it is very similar to eliciting intuitions" (##7) I think not, but the procedure had other serious problems (2015c: 51–53) and I abandoned it. I think that the experiments cited in the text do not have those problems and are different from eliciting intuitions. Sterelny is quite right though that these experiments do not discriminate "between causal descriptive theory and causal theories of reference" (##n. 6). We don't need experiments to reject causal descriptivism.

[76] See "Clarified Narrator's Perspective" in Sytsma and Livengood 2011; "reverse-translation probes" in Sytsma et al. 2015; "Award Winner Gödel Case" and "Clarified Award Winner Gödel Case" in Machery et al. 2015.

> misunderstood the question. (If we ask whether it rained at Trump's inaugural, we are not asking whether Trump, or anyone else, *thinks* it rained.) (Devitt and Porot 2018: 1554-1555).

The most telling count against folk intuitions comes, of course, from the recent tests of usage. These tests provide powerful evidence for antidescriptivism. So, for the folk intuitions to be reliable, they would have to be antidescriptivist. But studies from MMNS on show that the folk intuitions vary greatly and are far from consistently antidescriptivist.

However, Sterelny's doubts about the referential intuitions of philosophers have not been confirmed experimentally. Given the response to *Naming and Necessity* in the literature, we can reasonably assume that philosophers, particularly philosophers of language, generally share Kripke's referential intuitions. And this assumption has received some experimental support: in an experiment on a Gödel case, the intuitions of semanticists and philosophers of language were decisively antidescriptivist (Machery 2012a).[77] The tests of usage show that these intuitions are right.

Aside from testing referential intuitions and testing by elicited production, one might do a truth-value judgment test. In such a test an experimenter asks subjects to assess the truth value of some statement about the vignette. Machery, Christopher Olivola, and Molly De Blanc (2009; "MOD") did just that in response to Martí's criticism that MMNS should have tested usage (2009). MOD claimed that their truth-value judgment test was a test of usage. Martí disagreed (2012). I have sided with MOD on this, arguing that the test is "a somewhat imperfect" test of usage, its imperfection lying "in the fact that *it primes a certain usage*" (Devitt and Porot 2018). Martí is not quite convinced:

> MOD's questions are not as close to actual use as Devitt and Porot suggest, for they still require that the subjects think about how another speaker is using a name and what *she* is referring to when she uses it, a reflection that, as I have argued, is theoretical, and hence it is not obvious that it provides direct evidence of disposition to use. (##11).

We argued that MOD's test does not require theoretical reflection from the subjects because of the "disquotational" property of the truth term (Devitt and Porot 2018: 1559-1561).

The results in seven of our eight tests of usage (elicited production and truth-value judgment) were significantly antidescriptivist (and we think that we explained away the exception well enough). Still Martí surmises "that some semantic revisionists would disagree" with our antidescriptivist conclusion on the ground that "a small portion of subjects … appear to use names descriptively" (##11). We think that what those few subjects do is better thought of as "noise".

[77] For more discussion of this experiment, see Devitt 2012b: 23–24.

19.5 Metaphysics

19.5.1 The Definition of "Scientific Realism" (Godfrey-Smith)

Peter Godfrey-Smith and I have similarly realist views about science, as one would expect from Sydney boys, but we have been arguing for years about how best to *define* that realism. Any definition faces an obvious problem from the start: realism's opponents are so various; as Godfrey-Smith puts it in "Scientific Realism and Epistemic Optimism", "a diverse family of views – verificationism, radical constructivism, milder views like van Fraassen's constructive empiricism (1980), and so on" (##9). Rejecting this diverse family leads to two different strands in realist thinking: first, what Godfrey-Smith nicely calls "a generalized *optimism* about science, especially current science"; second, a "metaphysical" strand, famously rejected by idealism, about "the mind-independence of (much of) the world" (##2). I call these two strands, "the existence dimension" and "the independence dimension" of realism.[78] They are somewhat strange bed-fellows and so combining them in one doctrine can seem "a bit artificial" (##4). Because of this, I contemplated separating them into two doctrines when writing *Realism and Truth* but decided against partly because of terminological conservativism and partly because it would make things too complicated.

Also, it is worth noting that the two strands are not as unrelated as they might seem. For, a traditional route to idealism presumed optimism about our knowledge of the familiar world of stones, trees, cats, and the like. The route started, in effect, with an epistemological theory, a theory about what we could and could not know. It went on to argue that if the familiar world had the mind-independent nature required by realism, then we could know nothing about it: that view is "the very root of Scepticism" (Berkeley 1710: Sect. 86). But – here's the optimism – we obviously do know about it. So the familiar world is not mind-independent.[79] (This exemplifies the sort of argument that "Putting Metaphysics First" is against; see Sect. 19.4.1.)

In any case, I settled on definitions of "Scientific Realism" and "Strong Scientific Realism" that combined the two strands. A recent version of the strong doctrine is:

> SSR: Most of the essential unobservables of well-established current scientific theories exist mind-independently and mostly have the properties attributed to them by science. (2005b: 70).

I have long had a worry about this: "Couldn't someone be a realist in an interesting sense and yet be sceptical of the contemporary science on which realism … is based?" (1991: 20). Godfrey-Smith puts the worry like this, "one can have a realist metaphysics and be pessimistic about science" (##4); he mentions Popper as an

---

[78] Godfrey-Smith emphasizes the importance of being careful about "mind-independence" (##10). I agree and hope that I have been (1991: 14–17, 246–258).

[79] As Fodor put it nicely (to Rey), "Idealism is the effort to buy knowledge by selling off metaphysics."

example (#3). My concern about this led me then, and later (1997: 303–304), to contemplate definitions to accommodate these “realists”. Godfrey-Smith also tried to accommodate them (2003). Still, I stuck with the likes of SSR for two reasons. (1) It seemed to me a bit paradoxical to call a doctrine “realism” that had no commitment to the existence of anything in particular. (2) It seemed to me unhelpful, given the enormous historical role of skepticism in the realism debate about the observable as well as unobservable world, to call a doctrine that does not confront skepticism “realism”.

Godfrey-Smith has another worry. Quantum mechanics raises the possibility of a well-established scientific theory telling us “something very metaphysically surprising” (##5), namely, that we should reject the independence dimension. I have, of course, noted this distressing possibility (1984: 122; 1991: 132; 2005b: 69 n. 2) but I don’t see it as posing a problem for the definition. In my view, realism, *however defined*, is “an overarching empirical hypotheses” (1984: 18; 1991: 20) and so ought to be open to scientific falsification.

Finally, noting the qualifications in SSR – “most”, “essential”, “well-established” – Godfrey-Smith makes two objections to my definition:

> (i) There may be diversity across cases with respect to the appropriate level of confidence.
> (ii) There may be hidden diversity in what is being *claimed* by the science, in a way Devitt’s formulation does not accommodate. (##6)

In thinking about these objections, it is important to keep firmly in mind the *point* of definitions like mine. Godfrey-Smith thinks that such “a rough and coarse-grained summary … might have a role in some discussions (discussions of the relation between science and religion, for example)” (##5) but not, apparently, in philosophical discussions, at least not in the one I am engaged in. But why not? The point of these definitions is to come up with a doctrine, admittedly vague, rough and ready (1984: 11, 23; 1991: 13, 24), that is nonetheless, on the one hand, clearly rejected by members of that “diverse family” that Godfrey-Smith mentions; and, on the other hand, clearly the subject of current debates like that over the “no-miracles argument”. SSR fits the bill.

Consider objection (i). Godfrey-Smith rightly notes that “we have different kinds of evidence, and different shortcomings and reasons for doubt, in different fields” (##6). As a result, he thinks that SSR and its ilk involve “an excessive sculpting and refining of something that can only ever be a very rough and ready summary” (##6). But what changes would Godfrey-Smith recommend that avoids this “kind of false rigor” *whilst still fitting the bill*?

Objection (ii) is based on the role of modeling in science and has two themes. The first concerns “approximation and idealization” (##7). SSR presumes, as Godfrey-Smith notes, that in science “we *refer to objects* and *attribute properties* to them” (##6) Yet “in many fields, science aims to develop models that are *good approximations*” (##7). So drawing conclusions about what entities science is really committed to is a subtler business than SSR presumes. But, again, given the point of SSR, I think that these subtleties can be ignored.

The second theme is "the primacy of *structure* rather than *entities*" (##70). Scientists often think that their models show not so much that certain entities exist but that "the structure specified by these models is usefully similar to the structure of the system we are trying to understand" (##7). This is a serious worry because it challenges SSR's *truth*! I am not convinced by this challenge and so want to hold to my definition. But clearly the realistically inclined philosopher who *is* convinced should fall back to a definition of *structural* realism, as John Worrall (1989), and perhaps Godfrey-Smith, already have.

So, for the purposes of SSR, I think it is in order to overlook the subtleties Godfrey-Smith mentions. This is not to say, of course, that it is in order to do so to serve other respectable philosophical purposes; for example, the purpose of determining precisely what a particular scientific theory tells us about reality.

19.5.2 Biological Essentialism (Godman and Papineau)

19.5.2.1 Introduction

What is it *to be* a member of a particular biological taxon? *In virtue of what* is an organism say a *Canis lupus*? What *makes* it one? I take these to be various ways to ask about the 'essence', 'nature', or 'identity' of a particular taxon. The consensus answer in the philosophy of biology, particularly for taxa that are species, is that the essence is not in any way intrinsic to the members but rather is wholly relational, particularly, historical. Thus, in their excellent introduction to the philosophy of biology, *Sex and Death*, Sterelny and Paul Griffiths have this to say: there is "close to a consensus in thinking that species are identified by their histories" (1999: 8); and "the essential properties that make a particular organism a platypus … are historical or relational" (1999: 186). Samir Okasha endorses the consensus describing it as follows: we "identify species in terms of evolutionary history … as particular chunks of the genealogical nexus" (2002: 200). Philosophers of biology like to emphasize just how different their historical essentialism is from the influential views of Kripke (1980) and Putnam (1975).

In "Resurrecting Biological Essentialism" (2008d), I rejected the consensus in arguing that there is an intrinsic *component* to a taxon's essence. I called that doctrine, "*Intrinsic Biological Essentialism*" (IBE).[80] Still I accepted that there was also an historical *component* to a taxon's essence and have recently argued for this component in "Historical Biological Essentialism" (2018). As indicated by the title of their interesting paper, "Species have Historical not Intrinsic Essences", Marion Godman and David Papineau ("G&P") argue for the consensus historical doctrine and reject my IBE.

[80] This article has been criticized by Matthew Barker (2010), Marc Ereshefsky (2010), Tim Lewens (2012), Sarah-Jane Leslie (2013), and Matthew Slater (2013), all of whom are part of the consensus. I have responded Forthcoming-c.

My concern was with the essence of taxa thought to be in any one of the Linnaean categories. However, the custom in discussions of biological essentialism is to consider only species. I think this is a mistake but I shall go along with it here.

"Resurrecting" presented a positive argument for IBE and criticized the historical consensus. The positive argument had two parts (2008d: 351–355), which I now summarize.

19.5.2.2 Summary of Argument for Intrinsic Biological Essentialism (IBE)

The first part concerned the biological generalizations about the phenotypic properties of species and other taxa; generalizations about what they look like, about what they eat, about where they live, about what they prey on and are prey to, about their signals, about their mating habits, and so on. I argued that these generalizations have explanations that advert to intrinsic components of essences. In presenting this argument, I emphasized Ernst Mayr's (1961) distinction between "proximal" and "ultimate" explanations. (I preferred Philip Kitcher's (1984) terms for this distinction, "structural" vs "historical", but G&P prefer Mayr's and so I will go with that.) The explanations that featured in my argument were proximal ones about the underlying developmental mechanisms in members of a taxon that make the generalizations true. In contrast, ultimate explanations tell us how members of the taxon evolved to have such mechanisms.

In the second, related, part of the earlier argument, I claimed that a taxon's intrinsic essence explains why it is explanatory for an organism to be in a certain taxon:

> the generalizations we have been discussing reflect the fact that it is *informative* to know that an organism is a member of a certain species or other taxon: these classifications are "information stores" (Sterelny and Griffiths 1999, p. 195). But being a member of a certain taxon is more than informative, it is *explanatory*. Matthen points out that "many biologists seem committed to the idea that something is striped *because* it is a tiger" (1998, p. 115). And so they should be: the fact that an individual organism is a tiger, an Indian rhino, an ivy plant, or whatever, explains a whole lot about its morphology, physiology, and behavior. (2008d: 352)

Why does it? Because the essential nature of a taxon, to be discovered by biologists, causes its members, in their environment, to have those phenotypic properties. What nature? I argued that if our concerns are proximal, so they are with a nature that causes a tiger's development into an organism with those properties, the nature must be intrinsic.

Sarah-Jane Leslie claims plausibly that the traditional argument for essentialism "makes critical use of intuitions" (2013: 109). As can be seen, my argument for IBE does not. It makes critical use of biological explanations. (I should have emphasized this in "Resurrecting".) G&P also emphasize explanation:

> Essential properties are properties that explain all the other shared properties. For any Kind C, there will be some central common feature E possessed by each C, a feature that gives rise to all the other properties F shared by the Kind. The essential property thereby explains why the Kind supports multiple generalizations. (##3).

So, we are *very much* in agreement on this methodologically significant point. Yet we end up with very different conclusions.

19.5.2.3 G&P on Alice and Artifacts

G&P think of species as "historical kinds" which they contrast with "eternal kinds", using terms they take from Ruth Millikan (1999, 2000). They rightly think that we can throw light on species by considering some other kinds. So they begin their argument by discussing the essence of a range of kinds that they think are also "historical". And they claim that I think that species are "eternal". I find these terms quite unhelpful and so will not argue about their application to any kind. My focus in discussing species and some of these other kinds will be *very simple*: Do these kinds, whether appropriately called "historical" or "eternal", have an intrinsic component to their essence? For *that* is what is at issue with IBE. Whether the essence of species *also* have an historical component is *not* at issue. The clear difference between us is that I think that essential intrinsic properties answer proximal questions whilst G&P do not; they think that essential historical ones answer those questions.

G&P begin their argument as follows:

> Consider all the different copies of *Alice in Wonderland*, including the paperback with a front page torn off on Marion's bookshelf, the hardback in David's study, and the many others in numerous libraries and book stores across the world. These instances all share their first word, their second word, … and so on to the end. They also share the same list of characters, the same plot, and the same locations. We thus have a wealth of generalizations of the form *All copies of Alice in Wonderland are F*. Copies of *Alice in Wonderland* form a Kind. But the common properties of this kind are certainly not explainable by any common physical essence…. Rather, all these instances are members of the same Kind because they are all copies of an original. Their shared features are all due to their common descent from the original version written by Lewis Carroll. It is purely this chain of reproduction, not any common intrinsic property, that explains the shared features. (##4).

This is where G&P introduce their talk of *copying*, which is an important sign of their approach to essentialism. And, not surprisingly, copying *does* have a place in discussing the essence of a "copy" of *Alice*. I shall set *Alice* aside for a moment. But here's a quick initial thought: talk of *x* being essentially a copy of *y* seems a very unpromising way to reject an intrinsic component to the essence of *x*. For, *to be* such a copy, *x* must share the intrinsic properties of *y*!

G&P follow their remarks about *Alice* with some about artifacts:

> Many artefacts are like literary works in this respect. Earlier we alluded to all the features common to Vauxhall Zafiras…. But here again the commonalities are not explained by some common intrinsic property. While the Zafiras do have many physical properties in common, none of these is distinguished as the source of all the other common features. Rather their many similarities stem from their all being made according to the same original blueprint. They are constituted as a Kind by their common historical source. (##4).

After mentioning some other examples, they sum up their view of historical kinds:

> all examples will involve three central ingredients: 1) the existence of a model, 2) new instances produced in interaction with the model or other past instances, 3) this interaction *causes* the new instances to resemble past instances. A chain of reproduction thus generates the relevant historical relations that ground and explain the Kind. (##5)

19.5.2.4 Implements

Now G&P's claim is about "many" artifacts not all artifacts. But it is helpful to start by considering the essence of artifacts in general. And the first thing to note is that among artifactual kinds only the typical "trade-marked" ones like the Vauxhall Zafira or the iPhone are *essentially* artifacts. Consider "generic" artifacts like cars or smartphones.[81] These are, of course, made by us and they are so complicated that it may seem as if they *have* to be made by us. This makes it harder to see what is essential to being one of those things. So, let us consider something much simpler: a paperweight. To be a paperweight an object must *have a certain function*, the function of securing loose papers with its weight. Paperweights often have that function because they are artifacts designed to have it. But they often get that function in a very different way: a perfectly natural object like a stone or a piece of driftwood becomes a paperweight *by being regularly used* to secure papers. So, whereas having a certain function is essential to being a paperweight, being an artifact is not. Similarly, being an artifact is not essential to being a doorstop, a hammer, a pencil, a chair, or even a car or a smartphone. Putnam once remarked that chairs might have grown on trees. So might cars and smartphones!

We need a word for these functional objects. I call them "implements" So what is essential to an object's being a particular kind of implement is having a certain function. An object has that function in virtue of two properties. First, the object's relation to us or to some other organism: a car was made by us for a certain *purpose* and a nest was made by a bird for a certain *purpose*; a paperweight found on a beach is *standardly used* by us for a certain *purpose*. So relations to organisms are essential to kinds that are implements. But, it is important to note, not *one* of the relations G&P pick out in discussing *Alice* and the Zafira are essential to generic implements: these implements need not be "copies", have "a chain of reproduction", have "a model", have a "common historical source", or be made "according to [an] original blueprint".

Importantly, the second property in virtue of which an object has the function of an implement is intrinsic. For, the object must have *any intrinsic property required to perform that function*; thus, a paperweight has to have an intrinsic constitution that enables it to secure loose papers. No matter how much we intended something to be a paperweight, it won't be one unless it has that constitution; thus, a feather could not be a paperweight. And there may be more to the intrinsic component. An implement kind may have essential intrinsic properties *beyond those necessary for its function*, properties that distinguish it from other implement kinds *with the same*

81 I draw on my 2005a: 155–156.

function: pencils and pens are both writing instruments but they have different intrinsic essences.

Why should we believe these essentialist claims? They are intuitively plausible, I think, but we can do better than that. As G&P point out, "essential properties are properties that explain all the other shared properties". So we should look to such explanations to support our essentialist claims here as with species. Consider two examples. Why are paperweights useful weapons? Because the essential function of a paperweight requires it to have intrinsic properties (which we could spell out) that make it a good weapon. Why is it easier to erase writing from a pencil than from a pen? Because of the essential intrinsic difference between pencils and pens (a difference we could spell out).[82]

Turn now to G&P's examples. First, the Zafira. This is a typical trade-marked implement. Unlike the generic car, it is indeed part of the essence of a Zafira that it comes from a common historical source and is made according to an original (at least implicit) blueprint.[83] So, in that respect, G&P have the essence of a Zafira right. But in all other respects, they have it wrong. First, but not important, it is not essential that a Zafira be a "copy", have "a chain of reproduction", or have a "model". Second, and *very* important, many properties of Zafiras *are* "explained by some common intrinsic property".

A Zafira is essentially a car and so its essence includes all the essential properties of cars. So it must have the function of a car. So, first, it must be appropriately related to our purposes. Second, it must have all the intrinsic properties essential to functioning as a car; for example, having an engine, brakes, and seats. Something without those intrinsic properties – for example, a paperweight or a smartphone – *could not* be a car. Third, just as a pencil has an intrinsic essence that distinguishes it from other writing implements, so too does a car have an intrinsic essence that distinguishes it from other vehicles: from a van, truck, bus, pram, golf buggy, … Furthermore, a Zafira is a *special kind* of car and so there is even more to its intrinsic essence than to that of a generic car. It has to have *the particular sort* of engine, brakes, and seats peculiar to a Zafira; for example, it has to have the special 7-seat arrangement that is a strong selling point.

How do I know all this? Once again, as with species, we should look to explanation not just intuition. (a) Zafiras, like cars in general, are useful for suburban shopping. Why? The explanation is to be found in the intrinsic properties we have just alluded to. And, we should note, the historical fact that Zafiras are made by Vauxhall is irrelevant to that explanation: it would make no difference if the car had been made by Ford or not made at all, just found. (b) According to the advertisement, Zafiras are versatile, easy to drive, comfortable for a family, and disabilities friendly. Suppose they are. Each of these properties of the Zafira would be explained by its

[82] This discussion of generic implements bears on Millikan's discussion of chairs (1999: 56; 2000: 21).

[83] I say that the Zafira is a "typical" trade-marked implement because there may be some atypical ones that are not made at all; e.g. found objects ingeniously marketed as ornamental paperweights.

intrinsic essential properties. Once again, the relation of Zafiras to Vauxhall is beside the point.

So far, then, the trade-marked Zafira has just the same sort of three-part essence as the generic car: a relation to our purposes and some intrinsic properties that together give the Zafira its essential function as a car: and some other intrinsic properties distinctive of the Zafira. But there is a bit more to the essence of Zafiras. I agreed with G&P that, unlike cars in general, Zafiras must have a "common historical source": they must be made by Vauxhall. Why must they? As we have noted this relational property of Zafiras is irrelevant to the explanations we have been considering so far. But, it is not irrelevant to some other explanations; it may, for example, be part of explanations of the reputation and desirability of Zafiras (think of the fact that an iPhone is made by Apple), of its repair record, and so on. But even with these explanations, essential intrinsic properties are likely to be central.

I also agreed that Vauxhall must have made Zafiras according to an original (at least implicit) blueprint. But this requirement brings together what we have already identified as essential *without adding anything new*. (1) We have just agreed that Zafiras must be made by Vauxhall. (2) To say that they must be made according to that original blueprint is just to say that they must be made with the properties specified by the blueprint. *And those properties include the earlier-mentioned essential intrinsic properties of the Zafira*, the ones that carry the burden of explaining such commonalities of Zafiras as its special 7-seat arrangement. In brief, to say that the essence of Zafiras is to be made by Vauxhall according to the original blueprint is not to deny that the essence has an intrinsic component; rather it is to *entail* that it has. The essence that G&P propose is up to its ears in intrinsic properties.[84]

In conclusion, the view that the commonalities of a Zafira can be explained by a purely relational essence is, as I said in "Resurrecting" about an analogous view of the commonalities of a species, "explanatorily hopeless" (2008d: 363).

The story for *Alice* is, as G&P say, much like that for the Zafira. But they have both stories wrong. It is essential to a copy of Alice that it is a copy of the original manuscript produced by Lewis Carroll just as it is essential to a Zafira that it was made from Vauxhall's blueprint. But intrinsic properties are essential in both cases. For something to be a copy of *Alice*, it has to be a *linguistic* item with the *semantic* properties of Lewis Carroll's manuscript. That is what it *is* to be a copy and anything without such properties is simply not a copy of *Alice*. And it is those intrinsic linguistic properties – the "first word, … second word, … list of characters, … plot,

[84] They later mention another artifact, the pound coin: "We don't think all the pound coins pressed from some mould must have some common inner essence to explain why they share their many other joint properties" (##9). On the contrary, a common inner essence is necessary to explain their behavior in ticket machines and many other properties. Of course, the *most important* properties of the coin are its function of being worth one pound, which it has because of its relation to the Bank of England. Still, there is an intrinsic component to its essence.

and locations" – that explain such commonalities as that copies of *Alice* show great insight into the difference between quantifiers and names.[85]

19.5.2.5 Species

After a brief discussion of some other kinds, G&P finally turn to biological essentialism (##10). Their argument against IBE to this point is the suggestion that species are like other kinds, ones they classify as "historical", in not having an intrinsic component to their essence. I think that they are right to look to other kinds for guidance. But what we should learn from studying these others, supported by my discussion of Zafiras and *Alice*, is that *it is unlikely that the essence of any explanatorily interesting kind is wholly relational*. (*Being Australian* is my favorite example of an explanatorily *un*interesting kind that is probably wholly relational (2008d: 346). In any case, in the end, the rejection of IBE demands arguments about species themselves not about other kinds. We need to be shown that the proximal explanations of species commonalities do not rest on intrinsic properties. So that is what we now look for.

G&P contrast their position on species with mine:

> Why do their members all share so many properties? As we have seen, one answer would be to assimilate species to eternal Kinds, as Devitt does, and appeal to the common genetic make-up intrinsic to each member. But an alternative would be to view species as historical Kinds, and attribute their shared properties to their common ancestry, with their genetic make-up simply being part of the species' copying mechanism. (##8).

My examples of the sorts of shared phenotypic properties that we are talking about include: "ivy plants grow toward the sunlight … polar bears have white fur … Indian rhinoceri have one horn and Africa rhinoceri have two" (2008d: 351). And my claim is that an essential intrinsic underlying property of the kind in question is central to the proximal explanation of these commonalities, to the explanation of why each of these organisms develop to have the property specified for its kind. That intrinsic component of the essence, together with the organism's environment, causes the organism to have the specified property (352).

G&P give two reasons for rejecting my view and for taking species to be "fundamentally historical". But before considering them I want to address objections that came up in a helpful, though puzzling, private exchange with Papineau. Here is my version of the exchange:

Objection Your "essence" may explain why members of a species *S* have phenotypic property *P* or *Q* in common but what we want from an essence is to explain the fact that they "display a great number of commonalities" (##7); we seek the common cause of a *tightly correlated group* of properties.

[85] My discussion of blueprints and copying bears on Millikan's discussion of reproduction and copying in discussing the nature of kinds (1999: 54–56).

Response I agree that we want to explain that *as well* but if, as I have argued, a partly intrinsic essence explains *P* and it explains *Q*, then it follows that it explains the tight correlation of these two properties in *S*!

Objection But, your "essence" is not a *common* cause of these properties. Rather it is a conjunction of separate causes, one for *P*, one for *Q*, and so on. The essence should be a *single* common cause.

Response Let us go along with this intuitive talk of a "single" cause, for the sake of argument. If this requirement were good, then neither *S* nor Zafiras nor *Alice* would *have* an essence. For, *as a matter of fact*, one underlying property of S's ancestors causes *P*, another, *Q*, and so on; one property of the blueprint is responsible for a Zafira's brakes, another, for its special 7-seat arrangement: some semantic properties of the original *Alice* lead to its copies having one admirable literary property, others, another. Clearly, the singularity requirement is no good.

Turn now to G&P's two reasons. The second of these concerns microbial kinds. I did not address such kinds at all in "Resurrecting" and I later acknowledged that IBE may not apply to them. (Forthcoming-c) So I shall set the second reason aside and attend only to the first reason, "non-zygotic inheritance". Does it show that IBE is wrong about the non-microbial biological world? I think not. Indeed, I'm puzzled that they think that it does.

G&P start their discussion of non-zygotic inheritance by pointing out that "the children of a skilled forager" may not inherit their skills through their genes but rather "in other ways – for instance, by her explicitly training them, or by their implicitly copying her tricks. In this sense, there is nothing at all problematic about the inheritance of acquired characteristics" (##8). Indeed, there is not, and such inheritance is common in nature, as G&P bring out nicely. *But this throws no doubt on IBE*. For, the training and copying are part of the environment's causal role in the development of phenotypic properties. As G&P note (##9), I emphasize the obvious fact that "explanations will make some appeal to the environment" as well as to intrinsic essences (2008d: 352).

At this point, I'm sorry to say, G&P seem to go right off the rails: "nothing requires characteristic traits shared by species members to depend on genetic inheritance *at all*" (##9). Surely they can't really mean "*at all*"?! But they do:

> Why do all tigers grow up the same, and different from zebras, even though tigers and zebras are subject to just the same environmental influences? What could explain that, except their shared genetic make-up? Well, the answer is that tigers and zebras aren't subject to just the same environmental influences. Tigers are raised by tigers, while zebras are

> raised by zebras, and many of their species-characteristic properties can be due to this in itself – without any assistance from their genes. (##9).

Without *any* assistance?! Let's not beat about the bush: this is A Big Mistake. If it weren't for their shared genetic make-up, no tigers would acquire *any* traits from interaction with their parents (or with anything else). If G&P were right, a zebra brought up by tigers would have all the traits that tiger cubs acquire from interaction with their parents. That is surely not so. Chimps brought up by humans famously fail to learn a human language. (Indeed, language acquisition is a good example of the combined action of genes and environment in acquiring a trait.) Young cuckoos don't grow up like their foster parents.

So I see nothing in non-zygotic inheritance that counts against IBE. Aside from that, what about all the other commonalities of species? What explains why Indian rhinos have one horn and Africa rhinos have two? Or why tigers and zebras have stripes? These are certainly not traits acquired from watching parents. No case has been presented against the view that an intrinsic essence provides the proximal explanation of these traits.

There is a puzzle about this discussion of non-zygotic inheritance. It is offered to support the historical alternative to IBE: the shared properties are to be explained by "their common ancestry". And, as we have noted, in the background discussion it is claimed – falsely, I have argued – that copies of *Alice* and Zafiras have wholly historical essences. Yet an historical essence plays no role in the explanations G&P offer in discussing this non-zygotic inheritance.

In conclusion, there is certainly a relational, sometimes historical, component to the essence of literary works, artifacts and implements. But there is also an intrinsic component which plays a central role in the explanation of commonalities. And the same goes for species. IBE stands untouched.

References

Almog, J. 2012. Referential uses and the foundations of direct reference. *In Almog and Leonardi*: 176–184.

Almog, J., and P. Leonardi, eds. 2012. *Having in mind: The philosophy of Keith Donnellan*. New York: Oxford University Press.

Almog, J., P. Nichols, and J. Pepp. 2015. A unified treatment of (pro-)nominals in ordinary English. In *On reference*, ed. A. Bianchi, 350–383. Oxford: Oxford University Press.

Anderson, J.R. 1980. *Cognitive psychology and its implications*. San Francisco: W.H. Freeman and Company.

Antony, L. 2008. Meta-linguistics: Methodology and ontology in Devitt's *Ignorance of language*. *Australasian Journal of Philosophy* 86: 643–656.

Barker, M.J. 2010. Species intrinsicalism. *Philosophy of Science* 77: 73–91.

Berkeley, G. 1710. *Principles of human knowledge*.

Bianchi, A. 2015. Reference and repetition. In *On reference*, ed. A. Bianchi, 93–107. Oxford: Oxford University Press.

Bianchi, A., and A. Bonanini. 2014. Is there room for reference borrowing in Donnellan's historical explanation theory? *Linguistics and Philosophy* 37: 175–203.

Block, N. 1986. Advertisement for a semantics for psychology. In *Midwest studies in philosophy*, Studies in the philosophy of mind, ed. P.A. French, T.E. Uehling Jr., and H.K. Wettstein, vol. 10, 615–678. Minneapolis: University of Minnesota Press.

Burge, T. 1986. Individualism and psychology. *Philosophical Review* 95: 3–45.

Cappelen, H. 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.

Capuano, A. 2018. In defense of Donnellan on proper names. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0077-6.

Collins, J. 2006. Between a rock and a hard place: A dialogue on the philosophy and methodology of generative linguistics. *Croatian Journal of Philosophy* 6: 469–503.

———. 2007. Review of Michael Devitt's *Ignorance of language*. *Mind* 116: 416–423.

———. 2008a. Knowledge of language redux. *Croatian Journal of Philosophy* 8: 3–43.

———. 2008b. A note on conventions and unvoiced syntax. *Croatian Journal of Philosophy* 8: 241–247.

Culbertson, J., and S. Gross. 2009. Are linguists better subjects? *British Journal for the Philosophy of Science* 60 (4): 721–736.

Deutsch, M. 2009. Experimental philosophy and the theory of reference. *Mind and Language* 24: 445–466.

Devitt, M. 1972. *The semantics of proper names: A causal theory*. Harvard PhD dissertation. Available at https://devitt.commons.gc.cuny.edu/.

———. 1974. Singular terms. *Journal of Philosophy* 71: 183–205.

———. 1981a. *Designation*. New York: Columbia University Press.

———. 1981b. Donnellan's distinction. In *Midwest studies in philosophy*, The foundations of analytic philosophy, ed. P.A. French, T.E. Uehling Jr., and H.K. Wettstein, vol. 6, 511–524. Minneapolis: University of Minnesota Press.

———. 1984. *Realism and truth*. Oxford: Basil Blackwell.

———. 1985. Critical notice of *The varieties of reference* by Gareth Evans. *Australasian Journal of Philosophy* 63: 216–232.

———. 1989a. Against direct reference. In *Midwest studies in philosophy, Vol. 13: Contemporary perspectives in the philosophy of language II*, ed. P.A. French, T.E. Uehling Jr., and H.K. Wettstein, 206–240. Notre Dame: University of Notre Dame Press.

———. 1989b. A narrow representational theory of the mind. In *Representation: Readings in the philosophy of psychological representation*, ed. S. Silvers, 369–402. Dordrecht: Kluwer Academic Publishers. Reprinted in W.G. Lycan (Ed.), *Mind and cognition: A reader* (pp. 371–398). Oxford: Basil Blackwell.

———. 1990. Meanings just ain't in the head. In *Method, reason and language: Essays in honour of Hilary Putnam*, ed. G. Boolos, 79–104. Cambridge: Cambridge University Press.

———. 1991. *Realism and truth*. 2nd ed. Oxford: Basil Blackwell.

———. 1996. *Coming to our senses: A naturalistic program for semantic localism*. Cambridge: Cambridge University Press.

———. 1997. *Afterword. In a reprint of Devitt 1991*, 302–345. Princeton: Princeton University Press.

———. 1998. Naturalism and the a priori. *Philosophical Studies* 92: 45–65. Reprinted in Devitt 2010c: 253–270.

———. 2001. A shocking idea about meaning. *Revue Internationale de Philosophie* 208: 449–472.

———. 2002. Meaning and use. *Philosophy and Phenomenological Research* 65: 106–121.

———. 2003. Linguistics is not psychology. In *Epistemology of language*, ed. A. Barber, 107–139. Oxford: Oxford University Press.

———. 2004. The case for referential descriptions. In *Descriptions and beyond*, ed. M. Reimer and A. Bezuidenhout, 280–305. Oxford: Clarendon Press.

———. 2005a. Rigid application. *Philosophical Studies* 125: 139–165.

———. 2005b. Scientific realism. In *The Oxford handbook of contemporary philosophy*, ed. F. Jackson and M. Smith, 767–791. Oxford: Oxford University Press. Reprinted in P. Greenough and M. Lynch (Eds.), *Truth and Realism* (pp. 100–124). Oxford: Oxford University Press, 2006. Reprinted with additional footnotes in Devitt 2010c: 67–98. (Citations are to Devitt 2010c.).

———. 2006a. *Ignorance of language*. Oxford: Clarendon Press.

———. 2006b. Defending *Ignorance of language*: Responses to the Dubrovnik papers. *Croatian Journal of Philosophy* 6: 571–606.

———. 2006c. Intuitions in linguistics. *British Journal for the Philosophy of Science* 57: 481–513.

———. 2006d. Intuitions. In *Ontology studies cuadernos de ontologia: Proceedings of VI international ontology congress (San Sebastian, 2004)*, ed. V.G. Pin, J.I. Galparaso, and G. Arrizabalaga. San Sebastian: Universidad del Pais Vasco. Reprinted in Devitt 2010c: 292–302.

———. 2006e. Responses to the Rijeka papers. *Croatian Journal of Philosophy* 6: 97–112.

———. 2007. *Dodging the arguments on the subject matter of grammars: A response to John Collins and Peter Slezak*. Online at http://devitt.commons.gc.cuny.edu/online_debates/.

———. 2008a. Explanation and reality in linguistics. *Croatian Journal of Philosophy* 8: 203–231.

———. 2008b. A response to Collins' note on conventions and unvoiced syntax. *Croatian Journal of Philosophy* 8: 249–255.

———. 2008c. Methodology in the philosophy of linguistics. *Australasian Journal of Philosophy* 86: 671–684.

———. 2008d. Resurrecting biological essentialism. *Philosophy of Science* 75: 344–382. Reprinted with additional footnotes in Devitt 2010c: 213–249.

———. 2009a. Psychological conception, psychological reality. *Croatian Journal of Philosophy* 9: 35–44.

———. 2009b. The Buenos Aires symposium on rigidity: Responses. *Análisis Filosófic* 29: 239–251.

———. 2010a. What 'intuitions' are linguistic evidence? *Erkenntnis* 73: 251–264.

———. 2010b. Linguistic intuitions revisited. *British Journal for the Philosophy of Science* 61: 833–865.

———. 2010c. *Putting metaphysics first: Essays on metaphysics and epistemology*. Oxford: Oxford University Press.

———. 2011a. Deference and the use theory. *ProtoSociology* 27: 196–211.

———. 2011b. Experimental semantics. *Philosophy and Phenomenological Research* 82: 418–435.

———. 2011c. Methodology and the nature of knowing how. *Journal of Philosophy* 108: 205–218.

———. 2011d. No place for the a priori. In *What place for the a priori?* ed. M.J. Shaffer and M.L. Veber, 9–32. Chicago/La Salle: Open Court Publishing Company. Reprinted in Devitt 2010c: 271–291.

———. 2012a. The role of intuitions. In *The Routledge companion to the philosophy of language*, ed. G. Russell and D.G. Fara, 554–565. New York: Routledge.

———. 2012b. Whither experimental semantics? *Theoria* 27: 5–36.

———. 2012c. Semantic epistemology: Response to Machery. *Theoria* 74: 229–233.

———. 2012d. Still against direct reference. In *Prospects for meaning*, ed. R. Schantz, 61–84. Berlin: Walter de Gruyter.

———. 2013a. The 'linguistic conception' of grammars. *Filozofia Nauki* 21: 5–14.

———. 2013b. Responding to a hatchet job: Ludlow's review of *Ignorance of language*. *Revista Discusiones Filosóficas* 14: 307–312.

———. 2013c. Three methodological flaws of linguistic pragmatism. In *What is said and what is not: The semantics/pragmatics interface*, ed. C. Penco and F. Domaneschi, 285–300. Stanford: CSLI Publications.

———. 2013d. What makes a property 'semantic'? In *Perspectives on pragmatics and philosophy*, ed. A. Capone, F. Lo Piparo, and M. Carapezza, 87–112. Cham, Springer.

———. 2014a. Linguistic intuitions are not 'the voice of competence'. In *Philosophical methodology: The armchair or the laboratory?* ed. M.C. Haug, 268–293. London: Routledge.

———. 2014b. Linguistic intuitions: In defense of 'ordinarism'. *European Journal of Analytic Philosophy* 10 (2): 7–20.
———. 2014c. Lest auld acquaintance be forgot. *Mind and Language* 29: 475–484.
———. 2015a. Relying on intuitions: Where Cappelen and Deutsch go wrong. *Inquiry* 58: 669–699.
———. 2015b. Should proper names still seem so problematic? In *On reference*, ed. A. Bianchi, 108–144. Oxford: Oxford University Press.
———. 2015c. Testing theories of reference. In *Advances in experimental philosophy of language*, ed. J. Haukioja, 31–63. London: Bloomsbury Academic.
———. 2018. Historical biological essentialism. *Studies in History and Philosophy of Biological and Biomedical Sciences* 71: 1–7.
———. 2020. Linguistic intuitions: A response to Gross and Rey. In *Linguistic intuitions: Evidence and method,* ed S. Schindler, A. Drożdżowicz, and K. Brøcker, 51-68. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198840558.003.0004
———. Forthcoming-a. Three mistakes about semantic intentions. In *Inquiries in philosophical pragmatics*, ed. F. Macagno and A. Capone. Cham: Springer.
———. Forthcoming-b. *Overlooking conventions: The trouble with linguistic Pragmatism*. Cham: Springer
———. Forthcoming-c. Defending intrinsic biological essentialism. *Philosophy of Science*.
Devitt, M., and N. Porot. 2018. The reference of proper names: Testing usage and intuitions. *Cognitive Science* 42: 1552–1585.
Devitt, M., and K. Sterelny. 1989. Linguistics: What's wrong with 'the right view'. In *Philosophical perspectives, 3: Philosophy of mind and action theory*, ed. J.E. Tomberlin, 497–531. Atascadero: Ridgeview Publishing Company.
———. 1999. *Language and reality: An introduction to the philosophy of language*. 2nd ed. Cambridge, MA: MIT Press. 1st edition 1987.
Dickie, I. 2011. How proper names refer. *Proceedings of the Aristotelian Society* 101: 43–78.
Domaneschi, F., M. Vignolo, and S. Di Paola. 2017. Testing the causal theory of reference. *Cognition* 161: 1–9.
Donnellan, K.S. 1970. Proper names and identifying descriptions. *Synthese* 21 (3-4): 335-358.
Ereshefsky, M. 2010. What's wrong with the new biological essentialism. *Philosophy of Science* 77: 674–685.
Evans, G. 1973. The causal theory of names. *Aristotelian Society. Supplementary Volumes* 47: 187–208.
———. 1982. *The varieties of reference*, ed. J. McDowell. Oxford: Oxford University Press.
Field, H. 1973. Theory change and the indeterminacy of reference. *Journal of Philosophy* 70: 462–481.
Fodor, J.A. 1975. *The language of thought*. New York: Thomas Y. Crowell.
———. 1980. Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences* 3: 63–73.
———. 1987. *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
Godfrey-Smith, P. 2003. *Theory and reality: An introduction to the philosophy of science*. Chicago: University of Chicago Press.
Grice, P. 1989. *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
Gross, S., and J. Culbertson. 2011. Revisited linguistic intuitions. *British Journal for the Philosophy of Science* 62 (3): 639–656.
Haegeman, L. 1994. *Introduction to government and binding theory*. 2nd ed. Oxford: Blackwell Publishers. 1st edition 1991.
Hawthorne, J., and D. Manley. 2012. *The reference book*. Oxford: Oxford University Press.
Horwich, P. 1998. *Meaning*. Oxford: Oxford University Press.
———. 2005. *Reflections on meaning*. Oxford: Oxford University Press.
Jackson, F. 1998. *From metaphysics to ethics*. Oxford: Clarendon Press.

Kaplan, D. 1989. Afterthoughts. In *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein, 565–614. Oxford: Oxford University Press.

———. 1990. Words. *Aristotelian Society Supplementary Volume* 64: 93–119.

———. 2012. An idea of Donnellan. In Almog and Leonardi: 122–175.

Kitcher, P. 1984. Species. *Philosophy of Science* 51: 308–333.

Kripke, S.A. 1979a. Speaker's reference and semantic reference. In *Contemporary perspectives in the philosophy of language*, ed. P.A. French, T.E. Uehling Jr., and H.K. Wettstein, 6–27. Minneapolis: University of Minnesota Press.

———. 1979b. A puzzle about belief. In *Meaning and use*, ed. A. Margalit, 239–283. Dordrecht: Reidel.

———. 1980. *Naming and necessity*. Cambridge, MA: Harvard University Press.

Kroon, F. 1987. Causal descriptivism. *Australasian Journal of Philosophy* 65: 1–17.

Leslie, S.-J. 2013. Essence and natural kinds: When science meets preschooler intuition. In *Oxford studies in epistemology*, ed. T.S. Gendler and J. Hawthorne, vol. 4, 108–165. Oxford: Oxford University Press.

Lewens, T. 2012. Species essence and explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 751–757.

Lewis, D. 1984. Putnam's paradox. *Australasian Journal of Philosophy* 62: 221–236.

Loar, B. 1981. *Mind and meaning*. Cambridge: Cambridge University Press.

———. 1982. Conceptual role and truth-conditions. *Notre Dame Journal of Formal Logic* 23: 272–283.

Longworth, G. 2009. Ignorance of linguistics. *Croatian Journal of Philosophy* 9: 21–34.

Ludlow, P. 2009. Review of Michael Devitt's *Ignorance of language*. *Philosophical Review* 118: 393–402.

Machery, E. 2012a. Expertise and intuitions about reference. *Theoria* 27 (1): 37–54.

———. 2012b. Semantic epistemology: A brief response to Devitt. *Theoria* 27 (2): 223–227.

Machery, E., and S.P. Stich. 2012. The role of experiments. In *The Routledge companion to the philosophy of language*, ed. G. Russell and D.G. Fara, 495–512. New York: Routledge.

Machery, E., R. Mallon, S. Nichols, and S.P. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92 (3): 1–12.

Machery, E., C.Y. Olivola, and M. de Blanc. 2009. Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis* 69 (4): 689–694.

Machery, E., R. Mallon, S. Nichols, and S.P. Stich. 2013. If folk intuitions vary, then what? *Philosophy and Phenomenological Research* 86 (3): 618–635.

Machery, E., J. Sytsma, and M. Deutsch. 2015. Speaker's reference and cross-cultural semantics. In *On reference*, ed. A. Bianchi, 62–76. Oxford: Oxford University Press.

Martí, G. 2009. Against semantic multi-culturalism. *Analysis* 69 (1): 42–48.

———. 2012. Empirical data and the theory of reference. In *Reference and referring: Topics in contemporary philosophy*, ed. W.P. Kabasenche, M. O'Rourke, and M.H. Slater, 62–76. Cambridge, MA: MIT Press.

———. 2015. Reference without cognition. In *On reference*, ed. A. Bianchi, 77–92. Oxford: Oxford University Press.

Matthen, M. 1998. Biological universals and the nature of fear. *Journal of Philosophy* 95: 105–132.

Mayr, E. 1961. Cause and effect in biology. *Science* 134: 1501–1506.

McGinn, C. 1982. The structure of content. In *Thought and object*, ed. A. Woodfield, 207–258. Oxford: Clarendon Press.

Millikan, R.G. 1999. Historical kinds and the special sciences. *Philosophical Studies* 95: 45–65.

———. 2000. *On clear and confused ideas: An essay about substance concepts*. Cambridge: Cambridge University Press.

Neale, S. 2004. This, that, and the other. In *Descriptions and beyond*, ed. M. Reimer and A. Bezuidenhout, 68–182. Oxford: Clarendon Press.

———. 2016. Silent reference. In *Meanings and other things: Essays in honor of Stephen Schiffer*, ed. G. Ostertag, 229–344. Oxford: Oxford University Press.

Neander, K. 2017. *A mark of the mental: In defense of informational teleosemantics*. Cambridge, MA: MIT Press.

Okasha, S. 2002. Darwinian metaphysics: Species and the question of essentialism. *Synthese* 131: 191–213.

Orlando, E. 2009. General term rigidity as identity of designation: Some comments on Devitt's criticisms. *Análisis Filosófic* 29: 201–218.

Pepp, J. 2018. What determines the reference of names? What determines the objects of thought. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0048-y.

Pietroski, P. 2008. Think of the children. *Australasian Journal of Philosophy* 86: 657–659.

Putnam, H. 1973. Meaning and reference. *Journal of Philosophy* 70: 699–711.

———. 1975. *Mind, language and reality: Philosophical papers, volume 2*. Cambridge: Cambridge University Press.

Quine, W.V. 1961. *From a logical point of view*. 2nd ed. Cambridge, MA: Harvard University Press. 1st edition 1953.

Quine, W.V., and J.S. Ullian. 1970. *The web of belief*. New York: Random House.

Reber, A.S. 2003. Implicit learning. In *Encyclopedia of cognitive science*, ed. L. Nadel, vol. 2, 486–491. London: Nature Publishing Group.

Rey, G. 2006a. The intentional inexistence of language – But not cars. In *Contemporary debates in cognitive science*, ed. R. Stainton, 237–255. Oxford: Blackwell Publishers.

———. 2006b. Conventions, intuitions and linguistic inexistents: A reply to Devitt. *Croatian Journal of Philosophy* 6: 549–569.

———. 2008. In defense of folieism: Replies to critics. *Croatian Journal of Philosophy* 8: 177–202.

———. 2014. The possibility of a naturalistic Cartesianism regarding intuitions and introspection. In *Philosophical methodology: The armchair or the laboratory?* ed. M.C. Haug, 243–267. London: Routledge.

———. Forthcoming-a. A defense of the voice of competence. In *Linguistic intuitions*, ed. S. Schindler. Oxford: Oxford University Press.

———. Forthcoming-b. *Representation of language: Foundational issues in a Chomskyan linguistics*. Oxford: Oxford University Press.

Richard, M. 1983. Direct reference and ascriptions of belief. *Journal of Philosophical Logic* 12: 425–452.

Schwartz, S.P. 2002. Kinds, general terms, and rigidity. *Philosophical Studies* 109: 265–277.

Searle, J.R. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Slater, M.H. 2013. *Are species real? An essay on the metaphysics of species*. New York: Palgrave Macmillan.

Slezak, P. 2007. *Linguistic explanation and 'psychological reality'*. Online at http://devitt.commons.gc.cuny.edu/online_debates/ (Parts of this paper were delivered at a symposium on linguistics and philosophy of language at the University of New South Wales in July 2007.)

———. 2009. Linguistic explanation and 'psychological reality'. *Croatian Journal of Philosophy* 9: 3–20.

Slobodchikoff, C.N. 2002. Cognition and communication in prairie dogs. In *The cognitive animal: Empirical and theoretical perspectives on animal cognition*, ed. M. Bekoff, C. Allen, and G.M. Burchardt, 257–264. Cambridge, MA: MIT Press.

Smith, B.C. 2006. Why we still need knowledge of language. *Croatian Journal of Philosophy* 6: 431–456.

Stanley, J., and T. Williamson. 2001. Knowing how. *Journal of Philosophy* 98: 411–444.

Sterelny, K. 2016. Deacon's challenge: From calls to words. *Topoi* 35 (1): 271–282.

Sterelny, K., and P. Griffiths. 1999. *Sex and death*. Chicago: University of Chicago Press.

Stich, S.P. 1983. *From folk psychology to cognitive science: The case against belief*. Cambridge, MA: MIT Press.

———. 1991. Narrow content meets fat syntax. In *Meaning in mind: Fodor and his critics*, ed. B. Loewer and G. Rey, 239–254. Oxford: Basil Blackwell.

Sullivan, A. 2010. Millian externalism. In *New essays on singular thought*, ed. R. Jeshion, 249–269. Oxford: Oxford University Press.

Sytsma, J., and J. Livengood. 2011. A new perspective concerning experiments on semantic intuitions. *Australasian Journal of Philosophy* 89 (2): 315–332.

Sytsma, J., J. Livengood, R. Sato, and M. Oguchi. 2015. Reference in the land of the rising sun: A cross-cultural study on the reference of proper names. *Review of Philosophy and Psychology* 6 (2): 213–230.

Thornton, R. 1995. Referentiality and wh-movement in child English: Juvenile *D-Link*uency. *Language Acquisition* 4 (2): 139–175.

Unger, P. 1983. The causal theory of reference. *Philosophical Studies* 43: 1–45.

Van Fraassen, B.C. 1980. *The scientific image*. Oxford: Clarendon Press.

Weinberg, J.M., C. Gonnerman, C. Buckner, and J. Alexander. 2010. Are philosophers expert intuiters? *Philosophical Psychology* 23 (3): 331–355.

White, S.L. 1982. Partial character and the language of thought. *Pacific Philosophical Quarterly* 63: 347–365.

Worrall, J. 1989. Structural realism: The best of both worlds? *Dialectica* 43: 99–124.

Wulfemeyer, J. 2017. Bound cognition. *Journal of Philosophical Research* 42: 1–26.

Zerbudis, E. 2009. The problem of extensional adequacy for Devitt's rigid appliers. *Análisis Filosófic* 29: 219–237.