

There it is

BENJ HELLIE*

December 31, 2010

This article concerns the relation between *direct realism* and *perceptual justification*. My method will be this: presupposing a direct-realist outlook, I will develop a semantical framework for characterizing perceptual justification. I understand a ‘direct realist outlook’ to involve a cluster of methodological attitudes, such as: comfort with offering contrasting treatments of ‘good cases’ of veridical perception and ‘bad cases’ of matching hallucination; a desire to explain the so-called ‘transparency’ of perception; and an aversion to theories which would require pretheoretic opinion to be significantly in the dark about the characters of perceptual states which are rationally influential.

Road map: section 1 advances a picture of the nature of rationality and rational explanation in which consciousness plays a central role and then identifies the place of perceptual justification within this picture; section 2 links the theory of rationality to the traditional concerns of formal semantic theories; section 3 advances a direct realist-friendly semantical theory of justification by veridical perception; section 4 argues that this story cannot be extended to accommodate justification by hallucination; and the final section extends the discussion to a treatment of illusion, the *de re*, seeing-as, and other related phenomena.

If the reader is looking for a punch line, it is perhaps this: according to the direct realist, when one is taken in by hallucination, one’s picture of the world is *defective*—incoherent, inconsistent, unsatisfiable. For when one is taken in, the picture encoded within one’s background presuppositions about one’s condition is, according to the direct realist, inconsistent with the picture one accepts in one’s perceptual state. As a result, an explanation of the impact of such delusive

*Thanks are due to David Chalmers, Tom Hurka, Geoffrey Lee, Joshua Schechter, Susanna Siegel, and Scott Sturgeon, and especially Agustín Rayo, Andrew Sepielli, Nico Silins, Jessica Wilson, and Seth Yalcin.

hallucination on the remainder of one's cognitive system cannot be an intentional or rationalizing explanation: a 'ceteris paribus' assumption of coherence implicit in such explanation goes unmet; such situations break the boundary walls around the domain of this explanatory project. Attempts to treat both delusive and 'good' cases on a par, using the philosopher's favorite tool of intentional psychology or its rigorous counterpart, formal semantics, must fail. For such cases, we cannot offer a theory of justification; we must content ourselves with exculpation.

In this paper I will speak of 'kinds of stream of consciousness' or 'kinds of experience'; please understand my use as entirely 'inflationary', or, at least, as sufficiently inflationary to capture anything that might reasonably be meant by 'consciousness'. Some have used the notion of 'phenomenal property', understood as applying to properties of the stream of consciousness which are both (a) natural determinates of *consciousness* and (b) narrow, in a way roughly akin to my use; I am not confident, however, that there are any narrow natural determinates of *consciousness*, so I do not presuppose this. Indeed, qua direct realist, I presuppose the contrary. For more detailed discussion of this issue see Hellie 2010.

1 Perceptual justification

This article concerns 'perceptual justification'. This initial section attempts to provide a sharp content to this notion.

In the first three subsections, I discuss the notion of justification in general; their main lesson is that the core notion of rationality is something like manifest coherence of the stream of consciousness. If so, the most basic interpretation of the claim that *A* justifies *B* means something close to: from the first-person perspective, *B* was required to maintain coherence of the stream of consciousness in light of *A*. The notion of 'from the first-person perspective' is not *exactly* the same as 'according to the subject', because from time to time the subject is mistaken in what is required to maintain coherence—so, as we shall see, the direct realist must say. However, a guiding assumption of our discussion will be that we should be exceedingly cautious in positing divergence between what *is* required and what

one *takes* to be required.

The final subsection locates within this broader framework the phenomenon of distinctively *perceptual* justification.

1.1 Justification from the first-person

A theory of perceptual justification concerns *justification*. But what is that? Well, observe that there is at least some nebulous link between the claim that *B* is justified by *A* and the claim that *B* is a reasonable response to *A*; that citing *A* explains *B* by showing how it is reasonable; that *A* provides a rationalizing explanation for *B*. So our story about what justification is will begin by elucidating the structure of rationalizing explanation.

Sam's job is 'quality-controlling the widgets': when a widget comes down the conveyor belt, she is supposed to sort it off to the right just if it is defective, otherwise off to the left. We see Sam sorting a widget off to the right. Why did she do it? What is the rationalizing explanation for her action?

Sam knows best, so let's see how things look from her perspective:

Things are, going by looking, *thus*; so: certainly, this widget coming along is red. Red widgets are defective; so: certainly, the widget coming along is defective. I'm quality-controlling the widgets; so: sort this one off to the right!

This narrative or 'discourse' (understood as a string of sentences) is a sort of record of Sam's stream of consciousness. By this I don't mean that Sam's stream of consciousness involves an inner monologue in English that 'sounds like' that narrative; rather, that the words in this narrative capture or package under a linguistic mode of presentation the course of events as they are for Sam.

This narrative involves three transitions, marked by 'so'. Prior to each transition, Sam remarks on various facts. After each transition, Sam's psychological condition updates in a certain way. After each of the first two, Sam enters a condition of certainty, forming a new belief, as it were. These transitions therefore

count as episodes of ‘theoretical reasoning’. After the third, Sam issues a command to herself, commencing a new action. This transition therefore counts as an episode of ‘practical reasoning’.

From Sam’s point of view, these transitions ‘make sense’: entering those states of certainty, and commencing that action, are ‘apt’ as responses to the facts registered.¹ They are her answers to why she underwent those transitions. In this sense, for Sam, the responses are *explained* by the facts registered. Sam’s narrative therefore can be cast in reverse order as a chain of explanations:

Sort this widget coming along off to the right! Why? Because I’m quality-controlling the widgets, and the widget coming along is defective. Why the latter? Because red widgets are defective, and this widget coming along is red. Why the latter? Because things are, going by looking, *thus*.

We want to know how these explanations work: to understand what is explained, what explains, and why.

To do so, we need to uncover the logical forms of these explanations. It will help in doing so to have in hand a general theory of explanation: here’s a sketch of such a theory. Explanations considered as speech-acts are acts of answering ‘why’-questions. ‘Why is this F G ?’, one asks; ‘because it is H ’, another replies. For the most part, the reply is true just when (i) this F is indeed H ; and (ii) normally an F which is H responds by being G —if an F is H , it normally responds by being G —where which F s are counted as normal is of course context-relative.² More generally, if the reply is something like ‘because that M is N ’, we can say that (ii) has a more intricate form—for instance: if an N M bears R to an F , then normally the F responds by being G .

I will say that (i) is the *Fact* in the explanation while (ii) is the *Law*.³ I will also say that the presupposed notion of normality is the *Ceteris Paribus Condition*.

¹Something like the idea that the stream of consciousness is structured by a relation of sense-making is central to the system of Varela (1991); see also Thompson 2007.

In his PhD dissertation, Richard Yetter Chappell advances the idea that he attributes to Brentano and Sidgwick that ‘fittingness’ is the fundamental normative notion.

²For a congenial approach to the semantics of ‘normally’, see Veltman 1996.

³My view is somewhat similar to the classical deductive-nomological view of explanation.

Statements of Law contain tacit places for what is *Presupposed*: perhaps various matters of fact that are not made explicit in the explanation but which can be varied without rendering *ceteris no longer paribus*.

Let us suppose that Sam's narrative contains in some way the Fact part of the explanation of her sorting the widget off to the right. Then we can extract three stages of explanation from Sam's narrative: an 'outer' stage of practical-rational explanation, a 'middle' stage of epistemic-rational explanation by reasoning, and an 'inner' stage of epistemic-rational explanation by perception. The outer stage concerns Sam's decision to *sort the widget off to the right* rather than doing something else; the middle stage concerns Sam's adoption of the belief that *the widget is defective*, rather than a belief with some other content. For the time being, we will focus on the outer and middle stages to uncover some general patterns, leaving the inner stage for the final subsection.

In the 'outer' stage, the explanation regiments as follows:

Explanandum

Sort this widget off to the right!;

Law

If Π and (I am R-sorting the widgets \wedge this widget coming along is defective), then (sort this widget off to the right!);

Fact

I am R-sorting the widgets \wedge this widget coming along is defective;

Presupposition

Π ;

According to that theory, an explanation is a sentence conjoining a sentence stating a general law and a sentence stating a matter of particular fact, where the law and the fact entail the sentence stating a matter of particular fact which is the explanans. My story differs in the following ways. (a) The link between explanans and explanandum is not syntactic deducibility but rather truth of an indicative conditional. (b) My theory is not restricted to 'ideal' explanations but rather is to apply to ordinary explanations: the notion of normality need not be fully cashed out for the explanation to be true.

CP Condition

‘Rationality’.

While in the ‘middle’ stage, the explanation regiments as follows:

Explanandum

\square (this widget coming along is defective);

Law

If Π and this widget coming along is red and red widgets are defective,
then \square (this widget coming along is defective);

Fact

This widget coming along is red and red widgets are defective;

Presupposition

Π ;

CP Condition

‘Rationality’.

I now want to answer a number of questions about these regimented explanations.

1. Our Law-statements are indicative conditionals with ordinary indicative sentences in the antecedents and somewhat strange looking sentences in the consequents. Why?

An indicative conditional can have as its consequent sentences with a range of ‘forces’: indicatives, imperatives, questions, epistemically modalized claims. But it must have an ordinary indicative sentence in its antecedent: ‘If close the door!, P ’ and ‘if certainly Q , P ’ make no sense.

Still, the reactions up for explanation present themselves in ways that are the contents of these strange-looking questions, as we will now discuss.

2. I represent the outer Explanandum with an *imperative* sentence. Why?

Here's the story in brief. I am fond of the Aristotelian idea that 'action is the conclusion of practical reasoning', so to the extent that Sam's narrative to herself is regarded as the self-presentation of a course of practical reasoning, its conclusion—the part following 'So:'—should be a self-presentation of an action. Moreover, it seems to me that the content of an imperative 'A!' directed at oneself is something like the presentation to oneself of the 'commitment' to *A*-ing one secures when one is, in fact, *A*-ing. This is why we see an imperative as the outer Explanandum.

I'm also fond of the idea that actions are both practical explananda and practical explanantes (Thompson 2008). Sam's ongoing action of doing her job of R-sorting the widgets is what motivates her to sort this widget off to the right (rather than ignoring it). This is why we see an action-avowal in the Fact part of the explanans. We don't see an imperative as the Fact part because an imperative can't occupy the antecedent of a conditional.

3. I represent the middle Explanandum using a box. What does this box mean, and what is it doing there?

Well, ' $\Box P$ ' means *Certainly, P*. I will say more toward the end of this list on how I am understanding self-ascription of belief. The reason the explanandum is *Certainly, this widget coming along is defective* rather than *this widget coming along is defective* is that the latter is not a psychological fact and so cannot be rationalized by anything: in particular, not by Sam's looking at the red color of the widget. The former, by contrast, is a perfect fit: just as action is the conclusion of practical reasoning, certainty is the conclusion of theoretical reasoning.

The appearance of ' $\Box P$ ' in the narrative represents Sam's transition to certainty that *P*: at the moment that sentence shows up, Sam becomes certain of this.⁴

⁴In unpublished work Andrew Sepielli stresses the importance of the box as we shall understand it in psychological explanation.

We might wonder how ‘ $\square P$ ’ can serve as Middle Explanandum when ‘ P ’ is the Outer Explanans. Answering this will constrain our theory of the meaning of the box; more below.

4. Can we generalize beyond the particular case of Sam? What in general will one advance as Explananda and Facts in a first-person narrative of the sort rationalizing action, belief, or whatever else can be rationalized?

Let us say that one’s *stream of consciousness* is the sequence of experiences (understood as token occurrences) one undergoes. Let us say that the *kind* of an experience (at a time or over an interval) is what in particular it is like for one to undergo that experience at that time or over that interval: the distinctive natural maximal determinate of the property *being an experience* it instantiates. Let me reiterate that I do not think that experiences are of purely phenomenal kinds; rather, I think that the kind of an experience is a mix of the ‘indicative’ and the ‘imperative’: a kind of experience is a kind of action qualified by a picture of the world presupposed in or guiding that action.

Then I claim that no indicative or imperative sentence can appear in Sam’s first-person narrative unless its content is manifest in the kind of experience Sam undergoes. If properly understood, this is almost trivial. Sam’s first-person narrative characterizes her take on things: her ongoing actions, her picture of the world. And what could her take on things amount to other than what things are like for her? Fix Sam’s stream of consciousness while varying whether P (if this is possible) and Sam’s first-person narrative won’t change; alter her stream of consciousness and her first-person narrative will change.

Understood correctly, the claim is trivial; but the claim is easy to misunderstand.

First, note that my claim is that no claim can appear in Sam’s narrative unless its content is *manifest* in her stream of consciousness. My vagueness here is deliberate. Sam is certain that P and abiding an imperative to A just if

both *that P* (simpliciter) and *that Sam is A-ing* are manifest in her stream of consciousness. Since it might not be that *P*, and it might be that Sam fails in her try at *A-ing*, clearly the notion of ‘manifestation’ is not factive. But since only a fact can serve as an explanans, not every manifest hypothesis Sam might cite as an explanans is a genuine explanation. The position therefore builds in a certain degree of externalism about explanantes: Sam can take it that the explanation of her belief is that *P* when the explanation is rather that $\square P$; Sam can take it that her action is explained by her *A-ing* when the explanation is rather that she is following an imperative to *A*.

Second, Sam’s narrative need not be present in, or even explicitly structuring, her stream of consciousness in order to be accurate concerning it. The content of the narrative can be packaged in non-linguistic form in the structure of the stream of consciousness: in a certain course of actions, for instance.

Obviously all these explanations occur against the background of Sam’s total belief state: they presuppose the remainder of what she takes herself to know. I am inclined to think that this ‘picture of the world’ must also be manifest in Sam’s stream of consciousness, perhaps as providing it with form rather than with content, for the same reasons for which her passing considerations must appear in the stream of consciousness. I develop a way of making this seem plausible in my *Hellie* in preparation.

5. If the contents of explanantes and explananda are elements of the stream of consciousness, then in what does the truth of a statement of Law consist?

I offer several styles of response.

- (a) I advance a number of evocative synonyms. One’s reactions are governed by Law just if they *fit*, or are *reasonable* or *apt* in light of, or *appropriate* to, the explanantes; just if a stream of consciousness containing the explanantes but lacking the reaction would be *less than fully coherent* (manifestly inconsistent or incomplete).

- (b) We can operationalize the notion of a law-governed stream of consciousness. Not too surprisingly, we can operationalize it phenomenologically. If I myself violated a law of the sort under consideration—doing Sam’s job and having her opinions about the widgets in front of her but strangely failing to command myself as she does, I would experience a sense of incoherence or uncanniness: what is going on with me?, I would wonder.

We could think of this operationalization as a generalization of the linguist’s operationalization of ‘*’, intended to represent a failure of the ‘grammaticality’ of a sentence or discourse. Fresh linguists are trained up to recognize the sense of uncanniness resulting from thinking with such sentences as ‘Sam and Bill saw himself’, a sense then honed finely to consider discourses involving more subtle barbarism.

Indeed, we could think of * as corresponding to a special case of violation of rational law: perhaps * applies to *S* just if my thinking with *S* is never law-governed.

- (c) We can think of our Laws as being something like what traditional attempts at ascertaining the ‘laws of thought’ have been after: such doctrines as subjective Bayesianism and the Humean theory of motivation are attempts at advancing such laws.

In making this comparison, we need not commit to any particular such system of law as providing the truth. Indeed, it may even be that all familiar attempts are too restrictive, and that the truth is significantly more abstract. If so, discovering the laws of thought might be unhelpful as a way to settle disputes—a traditional goal of this project—but this would not impugn the truth of such laws.

- (d) We could think of being governed by law as somewhat akin to following a rule one accepts. The law might be more abstract than the particular rule, but perhaps the rule could be factored into its specific content and a much more general rule (‘under circumstances to which by one’s lights *A*-ing is an apt response, *A*!’), corresponding to one of

our laws. The analogy is this: just as following a rule is sometimes thought to ‘simply motivate’ transitions from one condition to another without further consideration of the need to follow the rule, being law-governed is simply being such as to respond in this or that way to these or those explanantes.

Do the Laws have the form of ‘objective’ or ‘subjective’ conditionals? Can their antecedents be statements of simple nonpsychological hypotheses, or must they always be statements of hypotheses about the psychological?

Well, action is typically required as that which explains in a case of practical reasoning, and action is in some sense ‘psychological’. At the same time, action is ‘external’, extending into the future and space outside the organism.

Moreover, on our story, it is evident that statements of nonpsychological hypotheses can serve as the antecedents of Laws: for example, the fact that this widget is red motivates Sam’s certainty that it is defective.

What if Sam is mistaken, though? What if the widget isn’t red? In that case, the fact that it is can’t explain Sam’s certainty that it is defective: there is no such fact to do any explaining. Sam’s certainty must rather have been motivated by something else: by her *certainty* that it was red.

Now, from the first-person perspective, P and certainty that P are equivalent. So motivation by P and motivation by certainty that P are also equivalent. So if we have Laws of rationality in hand, we can generate their first-person equivalents, which we might call ‘secondary laws’, where in each case an occurrence of P in the antecedent of a Law is replaced with an occurrence of $\square P$ in the antecedent of a secondary law. Perhaps if our aim is to distinguish between simple error and something more dramatic, full-blown irrationality, we can say that the latter is the failure to follow even a secondary law. If our point is understanding rational explanation, however, the move to secondary laws is not especially pressing. As we will see, from the first-person, P and

□*P* are equivalent. The fundamental variety of rational explanation is first-person explanation, so if we are interested in what is fundamental we are free to stick with primary laws.

Full-blown irrationality is worse. Just as Sam can be mistaken about what the explanation of her response is, Sam can also be mistaken about the explanatory structure of her stream of consciousness. When *ceteris* is not *paribus*, a stream of consciousness of a kind which should be followed with a stream of action of a certain other kind may be followed with a stream of consciousness of another kind. In general, we avoid engaging in that which is manifestly incoherent with what has gone before; so plausibly, one way for *ceteris* not to be *paribus* is for one to simply be insensitive to the manifest incoherence of one's situation. (Another might be for one to retain this sensitivity but for it to be rendered ineffective in guiding one's actions: in this case one might be faced with an uncanny sense of guidance by factors outside of oneself.)

6. What is this notion of 'rationality' mentioned in the *ceteris paribus* condition?

I don't intend anything heavyweight by this: I certainly do not mean 'ideal' rationality or even a manner of thinking or doing things that is liable to result in outcomes I value or beliefs of which I would approve.

Rather, rationality is a very permissive condition: more permissive than operation within the limits of behavior acceptable to this or that culture; more permissive than operation within the limits of human psychology. Rationality is a condition within which agents, whether human or animal (or robotic), typically operate. Still, it is less permissive than the notions of normality employed in biological explanation (if all agents are animals), or at least in chemical explanation.

We seem to grasp our notion implicitly in our grasp of rationalizing explanation. For instance, the 'sanity defense' in legal cases seems to reflect our sense that sometimes behavior that appears agentic if viewed with a squint

is not in fact agentic, but instead explicable solely in nonpsychological terms. (However, the legal notion of insanity may correspond more closely to that of human rationality.)

7. Does this story respect the ‘transparency of consciousness’?

Yes. The only claims one will ever offer as explanantes in the ordinary course of a first-person narrative are simple nonmental facts (e.g., that red widgets are defective) or facts about actions underway (e.g., that one is R-sorting the widgets).

Moreover, explananda, though explicitly concerning one’s own reaction to the nonmental (as one sees it) are very closely tied to such explanantes—in ways we will discuss under the next point.

However, there is also a sense in which consciousness is not entirely transparent. We will return to this point in two subsections.

8. I note that the middle explanandum is \square (this widget coming along is defective), while the corresponding outer explanans is *this widget coming along is defective*. Similarly, the outer explanans is (sort this widget off to the right)! But we can imagine the action one then commences, of sorting the widget off to the right, as an explanans in regard to some further action: e.g., picking up the widget gingerly. How are the indicative and the modalized nonmental claims related? And how are the indicative action avowal and the imperative related?

I want to advance a relation of *first-person equivalence* born to one another by the members of these pairs. This notion comes out in reflection on Moore’s paradox. The claims ‘*P*’, ‘certainly, *P*’, ‘I believe that *P*’, ‘it’s not the case that it might be that $\neg P$ ’, and ‘I know that *P*’ all bear this relation to one another; as do ‘I am trying to *A*’ and ‘I might *A*’. But none of these claims bear it to one another: ‘*P*’, ‘Bill believes that *P*’, and ‘Bill knows that *P*’; and nor do ‘Bill is trying to *A*’ and ‘Bill might *A*’.⁵

⁵The style I favor of treatment of the first-person avowals embroiled in Moore’s paradox first

Our notion of first-person equivalence can be elucidated in this way: two claims are first-person equivalent just if in advancing one but failing to advance the other, one is less than fully coherent. ‘That widget is green’, one asserts. But then one continues: ‘is it certainly green? I do not know’. Here one’s position is difficult to make sense of. Similarly for the converse: ‘certainly, that widget is green—but is it green? I do not know’. In this sense, one’s position puts one in this respect outside of the bounds of the rational. This need not be as radical as it sounds: one’s position is complex, and further aspects of it may be within the bounds of the rational; and in any event the failure of rationality may not persist for very long.

One sort of advancement is assertion or affirmation. Another is the issuance of an imperative. In issuing a command to open the window, one manifests one’s sense that the window is to be opened by the audience of the command. When the command is issued in public, the audience for the command is distinct from the issuer. If the audience accepts the command, the audience will also share one’s sense that the window is to be opened by them, and then go about opening it in confidence. Of course, however, the audience might fail to accept the command, and fail to share this sense. In this case, the command might fail to result in anyone’s even trying to open the window. Less dramatically, the audience might fail to completely accept the command, seeing some possibility of failure; if being cooperative, then, the audience might regard itself as merely trying to open the window.

However, when one issues a command to oneself, this cannot happen (short of one ‘part of the soul’ issuing the command to another). The issuer and the audience are one and the same, and thereby share the sense that they are to open the window, and therefore open it (if things go well). For this reason, issuing an imperative ‘A!’ to oneself results in going about opening the window in confidence. So it is not coherent to accept a self-command

appears in the philosophical literature, to my knowledge, in Gillies 2001. Gillies does not completely address the paradoxical aspect of Moore’s paradox, the *asymmetry* between ‘Bill believes that *P*’ and ‘I believe that *P*’, however, because he does not provide a theory of the meaning of the third-person ascription.

‘A!’ while failing to regard oneself as *A-ing*.

Conversely, if one regards oneself as *A-ing*, where this is an action rather than a mere rationally inexplicable process, one does so out of the sense that one is to *A*; and this sense results in acceptance of an imperative ‘A!’. So it is not coherent to regard oneself as *A-ing* while failing to accept the self-command ‘A!’.

We see then the first-person equivalence between the simple non-psychological claim *P* and the avowal of certainty $\square P$; and we see the first-person equivalence between the action self-ascription *I am A-ing* and the self-command ‘A!’.

This sets up the following structure of first-person narrative: simple non-psychological facts collectively explain transitions to states of certainty; which are first-person equivalent to simple non-psychological facts; which in turn are available for explanation of further transitions to states of certainty. In combination with action self-ascriptions, simple non-psychological facts explain the issuance of self-commands; which are first-person equivalent to action self-ascriptions; which in turn are available for explanation of further issuances of self-commands.

9. Our general theory of explanations suggests that the Fact explains the Explanandum by ‘bringing it under’ the Law. Mystery concerning the particular is dispelled when brought under the normal; and the normal is the general. But what is general in our Laws?

To superficial examination, they do not seem to be general: there is no argument place for a subject or time that is quantified over.

However, generality can enter in another way. The self-commander, and the avower of certainty, is the ‘transcendental self’ (or timeslice of such): not an object in anyone’s world, but a ‘limit’ of the world of each. Generality is reintroduced through the fictional shifting of perspective to other times and subjects: about which more is to come.

If this is all correct, then implicit in acting or updating one's picture of the world out of reason is the possibility of taking up the point of view of other times and subjects. If this seems too cognitively sophisticated to attribute to animals beyond humans and those we regard as 'higher', consider this: action is directed toward the future; and any animal recognizes the possibility of cooperation with others of its kind for the purposes of reproduction, and of competition with others of other kinds as either predator or prey.

1.2 Justification from the third-person

That was our story about the sorts of justifications—first-person rational explanations involving a manifest making sense of one's own psychological transitions—that give structure to Sam's stream of consciousness. What to do if *I* want to explain Sam's action or belief?

My task would start with my doing more or less what I did in the previous discussion: I 'push into', 'take up', 'project myself into', or 'simulate' Sam's point of view, and rehearse the narrative she advanced to herself.⁶

But this act of simulation can't be the end of the story. For suppose that Bill performs such an act. Then, here is how things are for Bill:

Here is how things are for Sam:

Things are, going by looking, *thus*; so: certainly, this widget coming along is red. Red widgets are defective; so: certainly, the widget coming along is defective. I'm quality-controlling the widgets; so: sort this one off to the right!

But at this point Bill is left with an undischarged push into Sam's point of view! Bill needs to pop back into his own point of view somehow. What should Bill do?

⁶My discussion here is highly influenced by Carnap's superb but rarely read paper Carnap 1932: the core sentence of this paper being 'The doctor here is his own diagnostic dog. ... The psychologist calls the behavior of the experimental subject 'understandable' ... , when his detector responds to it' (185). A somewhat similar perspective on psychological explanation would later be developed by Heal (2003) and others, though Carnap's characteristic epistemological and semantical sophistication would fail still to have penetrated this literature as I write this.

The problem, and its solution, may become a bit easier to see if we approach it through an analogous problem. Suppose I am explaining the narrative of *Gravity's Rainbow*. I give a speech like this:

... and then Slothrop participates in a Teutonic Springtime festival, in the role of the pig Plechazunga. The festival is broken up by American troops, and Slothrop goes on the lam again, wandering the countryside in the costume and the company of a pig ...

At the start of the speech I have pushed into the world of *Gravity's Rainbow*. At some point I will need to pop back out. And if I wanted anyone to learn something from my exercise of running through this narrative upon popping back out, I will need to come up with some way of bringing the narrative back with me: its content transformed, of course, because none of it is true. (Fictional truth isn't truth, and truth about a fiction isn't fictional truth.)

So if my aim in giving my little speech was to communicate some truths, I and my audience must be using some sort of tacit 'supposition discharge' schema whereby, by the speech act of discoursing through Δ within a push into the fiction of *Gravity's Rainbow*, one is licensed in certainty that $\Gamma(\Delta)$, upon having popped back out of the fiction. What is $\Gamma(\Delta)$? In general terms, it extracts from a discourse inside the fiction of *Gravity's Rainbow* a truth about the novel *Gravity's Rainbow*. More specifically, $\Gamma(\Delta)$ might be something metalinguistic like *in copies of Gravity's Rainbow are some sentences which imply the sentences Δ* . Or it might be something metapsychological like *the reader of Gravity's Rainbow can expect to enjoy imaginative immersion in a world in which Δ are true*.

We can express the logic of the situation with a balanced pair of introduction and elimination rules:

$$\text{In } GR: \Delta \quad \dashv\!\!\dashv \quad \square\Gamma(\Delta).$$

Here, ' $A \dashv\!\!\dashv B$ ' is a relation between kinds of 'thought act' to the effect that one who (implicitly) performs an A is not fully coherent unless one also (implicitly) performs a B . What these inference rules say is that anyone who, having pushed into the fiction of *Gravity's Rainbow*, presents the discourse Δ , must, to remain

coherent, be certain upon popping back out that $\Gamma(\Delta)$; and conversely, that anyone who is certain that $\Gamma(\Delta)$ thereby (implicitly) appreciates within the fiction of *Gravity's Rainbow* the discourse Δ .

Now back to third-person psychological explanation. Bill's recounting of Sam's first-person narrative isn't in itself informative either (except insofar as Bill trusts Sam in various opinions and has antecedently resolved to 'bring them back': the point is that it is not informative *about Sam*). (In saying that pushes into another stream of consciousness are not informative, this is not to say that they have no intrinsic value: such exercises may leave significant affective traces.) In order to explain how third-person understanding can be informative *about the other*, we need an operator-functor pair that stand to one another in roughly the same relation as the operator 'In $GR : \Delta$ ' does to the functor ' $\Gamma(\Delta)$ ': a 'thought-act' operator representing a push into someone else's point of view, and a discourse functor mapping discourses within the push to propositions about one's own certainty back after the pop. It will of course be helpful to have a bindable argument position for subjects as well so that we do not need to coin discourse operators anew for each new subject. So let's notate these objects as follows: 'FROM[X] : Δ ' is our thought-act operator, while ' $TO_X(\Delta)$ ' is our discourse functor.

The following inference rules relate them:

$$\text{FROM}[X] : \Delta \dashv\vdash \square TO_X(\Delta).$$

What these inference rules say is that anyone who, having pushed into X 's perspective, presents the discourse Δ , must, to remain coherent, be certain upon popping back out that $TO_X(\Delta)$; and conversely, that anyone who is certain that $TO_X(\Delta)$ thereby (implicitly) recognizes within X 's perspective that it is structured by the discourse Δ .

The left-to-right direction requires that there is no point of view one can take up that cannot be characterized in terms one regards as 'objective'. The right-to-left direction requires that if two subjects are indiscriminable in terms one regards as 'objective', one finds the same thing in taking up their points of view. The latter seems to be incompatible with the conceivability of zombies, if zombies

are regarded as beings objectively like us but lacking any stream of consciousness. It is only incompatible with the conceivability of immaterialism, however, if the objective is exhausted by the physical. While a venerable tradition supports this latter claim (‘the physical language is universal and inter-subjective’: Carnap 1932, 166), others challenge it (‘phenomenal information does not disappear from the third-person viewpoint, so it is not locating information (it is epistemically objective information), and so not the sort of thing that can be built into the center of a world’: Chalmers 2004, sec. 3): as for me, I am inclined to think that since colors are on things, it is unlikely that the objective is exhausted by the physical, but that is a side issue.

To get back to Bill, he can wrap up his understanding of Sam as follows:

Here is how things are for Sam—in the jargon, FROM[Sam] :

Things are, going by looking, *thus*; so: certainly, this widget coming along is red. Red widgets are defective; so: certainly, the widget coming along is defective. I’m quality-controlling the widgets; so: sort this one off to the right!

So: $\square \text{TO}_{\text{Sam}}(\Sigma)$;

Where ‘ Σ ’ abbreviates Sam’s speech.

What does one learn when one pops back to one’s own point of view? What does TO mean? To *which* proposition does TO map an arbitrary subject and discourse?

To make the question more tractable, I’m just going to assume that the explanatory structure of Δ is isomorphic to that of $\text{TO}_X(\Delta)$: if during the push into X ’s point of view, P explains Q , then after the pop back out, $\text{TO}_X(P)$ explains $\text{TO}_X(Q)$. It may be that the *style* of explanation on offer differs: that the popped-out explanations no longer look like rationalizations. This will depend on what the explananda and explanantes in the popped-out explanations are: what, in general, $\text{TO}_X(Q)$ is.

For concreteness, let X be Sam and Q be the sentence ‘that widget is red’. Then, one might think that $\text{TO}_{\text{Sam}}(\textit{that widget is red})$ is the proposition that *Sam*

believes that that widget is red. But while this is true, it is not helpful. After all, in saying that Sam believes that P , I say of Sam the same thing I say of myself when I say I believe that P . If the notion of belief here is supposed to be the same notion as the one implicated in first-person explanation, it generates Moore's paradox. In that case, 'I believe that P ' is in part an avowal of certainty that P rather than an 'objective' explicitly psychological self-predication. So the logical form of 'I believe that P ' is more like ' $\Box P$ '; and this lacks an argument place for a subject. So in order for the logical form of 'Sam believes that P ' to contain an argument place for Sam, it needs to involve something like ' $\Box P$ ' within the scope of TO. If so, then the logical form of 'Sam believes that P ' is ' $\text{TO}_{\text{Sam}}(\Box P)$ '—which of course contains an uneliminated occurrence of TO. So if *Sam believes that P* is a proposition, the logical form of the proposition contains an occurrence of TO: and in this case, the proposal has merely moved the bump in the carpet.⁷

But if not this, what? Answering this question strikes me as likely a messy enterprise. We need to know (a) what constrains the meaning of TO, and (b) what the objective facts are. Concerning (a), one thing we know about TO is its relation to FROM, as given in our inference rule. Our objective view on psychology therefore involves a sort of isomorphism between the objective and our best practices of simulation. Another source of constraint on TO is its use in interpretation: we observe such-and-such objective facts about X , and take these as motivating $\text{FROM}[X] : P$; in this case, we are perhaps treating those objective facts as default entailing $\text{TO}_X(P)$. Concerning (b), we might well promote some of the usual suspects to salience: the behavioral, the cognitive, the neurological, the physiological, the ecological. So in general $\text{TO}_X(P)$ will be a proposition about the behavioral (or cognitive, or neurological, . . .) that is, roughly, default entailed by facts which motivate $\text{FROM}[X] : P$ and/or an isomorph of best practices of simulation. Presumably what we want out of an isomorph may differ from case to case; if so, TO may be deployed adventitiously, having only a purpose-relative meaning. For present purposes, the answer to (a) is of little importance, and the

⁷'Where in the world is a metaphysical subject to be found? You will say that this is exactly like the case of the eye and the visual field. But really you do not see the eye. And nothing in the visual field allows you to infer that it is seen by an eye'—*Tractatus* 5.633.

exact answer to (b) is of no importance. What is important is that whatever proposition $TO_{\text{Sam}}(P)$ is, it concerns Sam in her objective aspects: and these, I venture, involve ‘in themselves’ nothing psychological.

If this is right, then we can say that popped-out explanations are not rationalizing explanations. My grasp on the notion of a rational norm comes, I think, solely from following the laws of rationality in guiding myself through my world. What I find to be rationalized are transitions to certainty (understood as a Moore-paradoxical notion), issuances of self-commands, and the like: nothing objective. Suppose the objective proposition to which $TO_{\text{Sam}}(P)$ is identical is the proposition that *Sam’s B_P -fibers are firing*. But that latter fact is not in the business of being rationalized, and although it can rationalize, it is not first-person equivalent to the fact that P , so it does not rationalize in the right way. A course of TO-facts therefore lacks the appearance of a rational stream. So when we move from FROM to TO the appearance of normativity drops away. Popped-out explanations may be biological or computational explanations, but they are not rational explanations.

This dissection of the logic of third-person psychological explanation is not much remarked upon in contemporary philosophy, though as I have remarked my story closely follows that of Carnap (1932). Indeed, my story is sharply at odds with the contemporary orthodoxy, consisting of something like functionalism about the propositional attitudes combined with either objective materialism or objective dualism about consciousness: an orthodoxy on which psychological properties show up from the objective perspective. Why has the line I am drawing been overlooked?

On my story, third-person psychological explanation inevitably involves us in a sort of double consciousness or aspect shift: via FROM, abandoning informativeness, we immerse ourselves in normativity; via TO, abandoning normativity, we learn more about which world is actual; we push back to normativity; then pop back to informativeness. We can imagine this push-popping accelerating; then eventually oscillating so rapidly that stages begin to blur together, to seem continuous; and we can imagine both streams running concurrently, in parallel,

tied together via our treatment of them as double aspects of a single underlying phenomenon.⁸ We might expect that, due to the intimacy with which they run in our thinking about others, the logical distinctness between the normative, subjective ‘pushed’ stream and the informative, objective ‘popped’ stream would be obscure, and the characteristic aspects of each stream would slop over into the other in the course of our theorizing about psychology.

1.3 Two consequences

This subsection draws out two consequences that will be of significance in our future discussion. I show a sense in which our framework predicts that consciousness is not transparent; and I argue that a theory of subjective justification is the only theory of justification we need.

The nontransparency of consciousness

In reading *Gravity’s Rainbow*, however immersive I may find the experience, my encounter with the affairs of Tyrone Slothrop, ‘Pirate’ Prentice, and the rest remains ‘within the scope’ of my ‘mainline’ stream of consciousness. I never for a moment regard myself as ‘really’ hearing the novel’s dialogue; the written text on the page never disappears—indeed, in light of the highly scholarly and hermeneutic approach demanded by that specific work, the written text insistently forces itself to the forefront of one’s stream of consciousness again and again.

Similarly, in engaging in an act of projection into the stream of consciousness of another, however compassionate I may become, the boundary between self and other never melts away, absorbing me into a more universal consciousness; I never manage to take even the slightest step taken toward Spiritual Unity, One Love, I and I. I never for a moment regard myself as identical to the object of sympathy, or as ‘really’ undergoing their experiences; the fact that what I am ‘really’ doing

⁸‘It is clear, however, that ‘A believes that *p*’, ‘A has the thought *p*’, and ‘A says *p*’ are of the form ‘*p* says *p*’: and this does not involve a correlation of a fact with an object, but rather the correlation of facts by means of the correlation of their objects’—*Tractatus* 5.542.

is sympathetically taking up the perspective of the other while remaining on a firm footing within my own perspective never disappears.

Perhaps souls more enlightened than this one can disintegrate the boundary between self and other. If so, this may be a capacity toward which we should all drive; but for present purposes, that is a side issue: the point is that in sympathetic projection, *much of the time*, the boundary remains firm.

If that is right, then in projecting into Sam's stream of consciousness, her stream of consciousness qua object of my projection becomes an *object of consciousness to me*. From my 'mainline' stream, I recognize Sam's stream of consciousness as that into which I am projecting. This is not the same attitude I ordinarily take toward my own mainline stream of consciousness: ordinarily, my mainline stream is *that from which* other entities are taken as objects. I both simultaneously inhabit and witness Sam's stream of consciousness. And as an object I witness, Sam's stream of consciousness is *not* transparent to me.

But now note that I can equally well take this attitude toward my own stream of consciousness. Instead of simply living through my present stream of consciousness, I can add to this an uncanny spectatorial perspective on it, in which I regard Hellie as an object like Sam into whose stream of consciousness I can project. In such a case, I both live through my stream of consciousness (a stream which involves an act of projection) and treat it as object. As that through which I live, it is transparent. But as an object I witness, it is *not* transparent.

Here is an example. I engage in a fair bit of thinking, some of it in words: so-called silent soliloquy. This thinking *uses* these words to, say, plan out the course of cooking a meal. But I can also take the spectatorial perspective on the words I am using in thought: these entities which form my stream of consciousness, the content of which is a developing picture of the future which will then guide my cooking activities, are transformed from form to content. An especially advanced stage of this spectatorial perspective is reached in the mastery of formal calculi. In engaging in mental arithmetic, the language with which I normally negotiate the world I manipulate becomes itself an object of manipulation: 'how many 'zeroes' did I have back in that calculation of the volume of the petroleum—nine or ten?',

I ask myself. Here the ‘zeroes’ of which I think are mental words. Confronted with a question, I push into ‘mention’ mode, shove digits around until I get what I wanted, and then pop back out to ‘use’ mode again.

‘Objective’ and ‘subjective’ ‘normativity’

The liquid in this glass is gasoline, though Fred thinks it is gin and tonic. If I were Fred, I would not drink the liquid in that glass.⁹

That case is pretty extraordinary: we might take away more representative lessons about practical reasoning from a more mundane case. Angela, engaging in a bit of urban tourism, is somewhat hungry. She finds herself in front of the local branch of Tasty Burrito, which dependably provides an OK lunch. What she doesn’t know, but could—like I did—learn by combing through the *Downtown Weekly* she is holding, is that across the street lies Taco Bueno Sabor, where very delicious lunches are served. If I were Angela, I would cross the street and dine at Taco Bueno Sabor.

That’s my story. Bill follows up with a different story:

Something else Angela—like Hellie, it seems—does not know (but could—like I did—learn by asking the guys hanging around on the corner) is that up the block a bit lies Taco Caliente Peligroso, where mindblowingly delicious lunches are served. If *I* were Angela, I would go up the block and dine at Taco Caliente Peligroso.

I can put Bill’s claim by saying that if *Bill* were Angela, Bill would go up the block and dine at Taco Caliente Peligroso.

Selena follows up with a different story still:

Unlike Hellie and Bill, I prefer the familiar flavors of Tasty Burrito to the more rustic offerings available elsewhere on the block. So if I were Angela, I would stay right here and dine at Tasty Burrito.

⁹Thanks to Seth Yalcin for calling my attention to the significance of ‘if I were you’ statements. Yalcin’s work in progress with Geoffrey Lee catalogues a range of support for the view that such statements differ substantially in their semantic properties from ordinary subjunctive conditionals.

If Selena were Angela, that is to say, Selena would dine right there at Tasty Burrito.

Now Mark follows up:

I care nothing for the pleasures of the flesh: I care only about alleviating suffering. If I were Angela, I would purchase that potato and gnaw on it to alleviate hunger, and send the money I saved to Oxfam.

If Mark were Angela, Mark would forego restaurants entirely.

We can hear also from Ellen:

If I were Angela, I would gnaw on that potato too, but not for Mark's reasons: he cares way too much about morality. Rather, because the scientific papers I have read prove that eating cooked food indoors shortens the lifespan by two decades. Well, I don't know. Angela might not be aware of these results or might be skeptical or might not care or might care but care more in the moment about the satisfactions of a hot meal. So I guess maybe if I were Angela, like Bill, I would go to Taco Caliente Peligroso. Well maybe I wouldn't. How would I know to do that? I'd have to ask those guys, and they might take that the wrong way. So maybe like Hellie I would go to Taco Bueno Sabor. But in order to do that I'd have to spend a bunch of time thumbing through *Downtown Weekly*, and I guess as Angela I'm hungry right now. So I guess I'd just go to Tasty Burrito. Or maybe I wouldn't even be facing this decision: rather than participating in urban tourism, I'd be in the office working on a paper. Well, maybe Mark is right: instead I'd be in the field nursing the wounded. But how would I get myself into a position of caring so much about morality? I'd have to spend a lot of time retraining myself into a different lifestyle and given the demands of the job that's time I don't have . . .

Ellen tells a rather more complex story than any of our other conversants.

Plausibly, claims like 'if S were T , S would A ' concern the doings of an agent with a stream of consciousness that is in some salient way a 'blend' of S 's and T 's: let $[S/T]_c$ be the agent blended in the way salient in context c . The logical form of the claim is something along the following lines: FROM $[[S/T]_c] : A!$. (Namely, 'Let's hear from an agent whose stream of consciousness is the salient sort of

blend of yours and mine: blah blah blah; so: A!'. Here 'blah blah blah' narrates the salient blend.) To the extent that 'if I were you, I would A' serves to motivate the audience to A, it does so perhaps by motivating the audience to become the blended agent, who then issues the predicted self-command 'A!'.

In blending agents, what is preserved, and what is held fixed? As our examples show, there seems to be almost no discipline to this practice: we can add or subtract information, accommodate habits of theoretical reasoning or not,¹⁰ accommodate the costs of acquiring information or not, flip around matters of personal taste or the relative weights of morality and self-concern, accommodate the costs of altering such matters of 'value' or not.

To the extent that notions like that of a 'reason' or a 'reasonable response' or what one 'ought' to do are understood along the lines suggested in the discussion of this section, as based in our subjective sense for a fit reaction to what is given in one's stream of consciousness, a notion of what I 'objectively ought' to do could only be understood as what some Very Special Agent would do, if they were me.¹¹ The apparent lack of discipline to our practice of blending agents suggests, to me at least, that there is probably no natural such Very Special Agent. The less liberally-minded, no doubt, will disagree.

1.4 *Perceptual justification*

To affirm my justification in a certain reaction is to say that there is some fact present within my stream of consciousness to which this reaction strikes me as appropriate; that which justifies my reaction is that fact present within my stream

¹⁰This factor is recognized in *The Lion in Winter*:

Henry II: Good God, woman, face the facts.

Eleanor of Aquitaine: Which ones? We have so many.

This concern seems to undermine the following case for the 'objectivity' of 'right' found in Ross 2002 and taken up more recently by Graham (2010): what Bill ought to do is what he would do if he knew everything—which, if he does not, might be different from what in Bill's actual view he ought to do. I find the hypothesis that Bill 'knows everything' to be incomprehensible.

¹¹When an ethical law of the form, 'Thou shalt ...' is laid down, one's first thought is, 'And what if I do not do it?'—*Tractatus* 6.422.

of consciousness to which my reaction strikes me as appropriate. If this is right, perceptual justification of belief is the presence in one's stream of consciousness of something 'perceptual' to which belief of a certain sort strikes one as appropriate.

To get back to our story about Sam, we observed a case of perceptual justification in the *inner* explanation: recall, 'why be certain this widget coming along is red? —Because things are, going by looking, *thus*'. This explanation regiments as follows:

Explanandum

□(this widget coming along is red);

Law

If Π and (things are, going by looking, *thus*), then □(this widget coming along is red);

Fact

Things are, going by looking, *thus*;

Presupposition

Π

CP Condition

'Rationality'.

The remainder of this paper will be occupied in explaining what this Fact means and why the Law is one.

Before launching into that, let us make more vivid what the presence of perceptual justification amounts to: we do so by contrasting our case in which it is present with cases in which it is absent. Sam's job, recall, is to sort the widgets that go by: defective ones off to the right, the rest off to the left. Sam sees a red one going by, and, knowing that red ones are defective, sorts it off to the right. A bit later, Sam sees a green one going by, and sorts it off to the left. Both of these *actions* were justified: were apt responses to the character of Sam's stream of

consciousness; so evidently Sam believed when the red widget was going by that the widget going by was red, and believed when the green widget was going by that the widget going by was not red. This variation in Sam's beliefs is an aspect of her stream of consciousness: after all, it justifies what are manifestly variable reactions.

This variation in belief is something that admits of justification and failure of justification. There is nothing about Sam's stream of consciousness that strikes her as in any way bizarre, uncanny, incoherent. By contrast, we can imagine subjects with the same sort of variation in belief for whom the variation is incoherent. Here are some examples:

- Claire performs the same task as Sam, and therefore manifests the same sort of variation in her beliefs as Sam. But Claire is clairvoyant (BonJour 1985): she cannot see, and all her other senses are otherwise just like Sam's. There is an aspect of Sam's stream of consciousness that does not correspond to any aspect of Claire's stream of consciousness. Claire's variation in belief is bizarre, incoherent: why does she believe in this way and act accordingly? This question raises itself as one with no apparent answer: by contrast, not so for Claire. It is plausible that the aspect missing from Claire's stream of consciousness but present in Sam's is what makes the difference here.
- Fidel performs the same task as Sam, and therefore manifests the same sort of variation in his beliefs as Sam. Fidel is not clairvoyant: he can see, and all his other senses are otherwise just like Sam's. But something bizarre happens with Fidel as the green widget goes by. Fidel becomes 'spectrally inverted': his visual system temporarily rewires itself so that it is in the same intrinsic condition as it was when the red widget was going by—not to be coy, the green widget looks red to Fidel. His variation in beliefs is *despite* this sameness in how the widgets look to Fidel, and is not explained by any background expectation or prior information to the effect that he would become inverted. As a result, his variation in belief about the colors of the widgets is bizarre, incoherent: why does he believe in this way and

react accordingly? The question raises itself and has no apparent answer; not so for Claire. It is plausible that the variable aspect in Sam's stream of consciousness which in Fidel's stream of consciousness is invariant explains this contrast.

To put things in theoretically loaded terms, Claire and Fidel vary in belief as does Sam, but Claire lacks any (relevant) *subjective perceptual state* over the period of variation, while Fidel's *subjective perceptual states* are constant in character (a constancy not matched by variation in background expectations); for this reason, Claire's and Fidel's streams of consciousness are incoherent while Sam's is coherent. Sam's beliefs cohere with her course of subjective perceptual states and are justified by them; Claire lacks any subjective perceptual state to do justificatory work; Fidel's subjective perceptual state is not of the right kind to do the specific sort of justificatory work required.

The job description of subjective perceptual states, then, is this: to serve as aspects of the stream of consciousness which provide justification of the sort of variation in belief of the sort observed in Sam's case but missing from Claire and Fidel's cases.

This job description is vague. 'Of the sort observed'—what sort is that? This vagueness is deliberate. We think of a clairvoyant as someone who shares our 'beliefs' but lacks perceptual states, and an invert as someone who shares our 'beliefs' despite having the wrong perceptual states, and is therefore in an uncanny position. But the notion of a belief is not, in my view, an especially clear one: significantly less clear, I think, than those of a perceptual state or of an action. If I see some object 'as a hammer' and therefore pick it up as a means to a job of driving nails, do I believe it is a hammer or do I go straight from the perceptual state to the action? I don't know the answer here. However, this paper is supposed to address the justification of belief by perception, and as I understand the belief–perception contrast the clairvoyant thought experiments are supposed to operationalize it.

Another question: how 'rich' is a perceptual state? Does it accommodate only what one is 'attending to' (looking at, feeling, and the like), or does it accommodate everything one *perceives*, a broader classification? My use concerns the

former: any residue has no direct presence within one's stream of consciousness and therefore cannot be rationally significant. Of course the residue can causally influence one's stream of consciousness, not least by 'capturing attention'; but in my view attention capture is a paradigm case of an arational update in the stream of consciousness.

Note also: I'm inclined to regard my remarks about perceptual states as applying without significant alteration to 'sensory' and 'mood' states. Obviously the affective contribution of this class of states is of considerable significance, but I will not be able to address this contribution here.

So my answer is this: if you have a coherently conceivable clairvoyant or invert thought experiment in hand, then my story concerns the distinction between your thought experiment and the normal course of things.

2 Justification and semantical theory

Roger Bacon teaches us that 'the things of this world cannot be made known without a knowledge of mathematics'; my aim in this article is to make perceptual justification better known by following Bacon's advice. There is a piece of mathematics that was developed for the purposes of understanding rationality in general: formal semantics. Frege invented truth-functional semantics in order to explain valid inference; Wittgenstein extended and generalized Frege's discoveries, inventing possible worlds semantics in order to characterize the limits of coherent thought. This article exploits possible worlds semantics in order to provide a Bacon-approved story about perceptual justification. But since everyone uses formal semantics differently, we will need to spend some time explaining how it is being used here: what the relevant mathematics looks like, and how it helps us understand rationality.¹²

To begin, I invite you to consider a certain situation, about which I will make a

¹²My story is very heavily influenced by conversation with Agustín Rayo, and by the theory developed in his typescript *Possibility and Content* (Rayo in preparation); and also by the *Tractatus* (Wittgenstein 1921/1974).

series of stipulations: take these as ‘monotonic’ in the sense that each stipulation characterizes the situation under consideration more determinately. OK, here’s the first one: in this situation, roses are red. Second, violets are blue. Third, Bertrand Russell lived exactly 35,689 days (for the record, this stipulation is actually true). Fourth, it did not snow in Toronto on 14 December 2010 (for the record, this stipulation is actually false: really, it did snow in Toronto on 14 December 2010). And, fifth and finally, it is not the case that Bertrand Russell lived exactly 35,689 days.

How did that go for you? I imagine something like this. The first stipulation was fine: you conjured up a picture of the world in which roses are red—an easy thing to do, since roses are red and indeed you are already aware of this. The second stipulation also created no special difficulties: you merely made the situation under consideration more determinate in a way that tracked added determinacy in your correct opinions about the world. The third stipulation probably was also unproblematic: while you were doubtless not aware that this stipulation was true, and had indeed probably never considered this stipulation, you had no difficulty contributing still further determinacy to the picture of the world under construction by adding this stipulation. The fourth stipulation should also have raised no problems: adding determinacy to a picture of the world by stipulating falsehoods (as well as stipulating truths of which one was unaware) is easy. But accommodating the fifth stipulation, I hope, raised the gravest of difficulties. The third stipulation required that Russell lived exactly 35,689 days; and since we assumed at the outset that our stipulations would contribute detail monotonically to the situation under consideration, when asked in the fifth stipulation to add the further detail that Russell did *not* live exactly 35,689 days, we do not know what to do. We do not understand how there could be a situation in which both Russell lived exactly 35,689 days and it is not the case that Russell lived exactly 35,689 days. Such a picture of the world cannot be coherently drawn up, in our view.

If such a scenario cannot be made sense of, it surely cannot be coherently affirmed as accurate to the actual course of events: supposition requires less of one than affirmation, not more. And if I cannot make sense of such a scenario or

coherently affirm it, nor can I make sense of someone else's doing so: any attempt to grasp their sense of the world from within would ground out on the fact that the picture of the world they affirm or are considering is incoherent.

If a certain pattern of affirmations, all bundled together into a single affirmation of a picture of the world more detailed than the picture affirmed in any one of them, would inevitably embroil someone in affirmation of an incoherent picture of the world in this sense, and therefore in incoherence, we can say that those affirmations are incoherent. For example, affirming that roses are red and together with this affirming that it is not the case that roses are red would embroil one in affirming that roses are red and it is not the case that roses are red. But then one would be affirming a picture of the world which is incoherent; and in that sense one would be oneself incoherent. So we may say that affirmation that roses are red is not compatible with affirmation that it is not the case that roses are red. Nor for that matter is it compatible with agnosticism that it is not the case that roses are red, or uncertainty whether this is so. Any attitude other than all-out rejection cannot be made sense of.

We observe a noteworthy parallelism here, which can be cast into a useful mathematical formalism with the help of the device of the 'possibility'. What is a possibility? Something like a possible world, or a centered possible world, or a cell in a 'chunky' or 'grainy' partition of the class of possible worlds or centered possible worlds.¹³ I will not agonize too much over the important formal differences among these various choices. In particular, throughout the paper, I will just ignore the central—and yet for present purposes, tangential—issue of diachronic change in *de se* attitudes, treating updating in effect along lines that are entirely 'de mundo'.

(Advance apology to the reader: sometimes I might say 'possible world' meaning 'possibility'. If so this is due to sloppy editing resulting from time pressure.)

We can think of the scenarios in which roses are red, relative to a given 'pri-

¹³If possibilities are grainy in this way, the status of *P* at a possibility is trivalent. For the most part I will ignore this complication.

mal' set of possibilities, as modeled by the subset of the primal set containing exactly those possibilities in which roses are red. So we can think of the stipulation that in a certain scenario, roses are red, as narrowing down the set of possibilities under consideration to those from the primal set in this set. We can think of the scenarios, relative to *this* set, in which violets are blue as modeled by the subset of this set containing exactly those possibilities in which violets are blue. And so forth. We are fine up through the fourth stipulation in the sense that some possibilities remain for consideration. But with the fifth stipulation, we can no longer find any possibilities from among the primal set meeting all our stipulations at once. Extracting from a set of possibilities in all of which Russell lived exactly 35,689 days the subset of all those possibilities in which it is not the case that Russell lived exactly 35,689 days leaves us with a null set of possibilities. (The attempt at updating one's preexisting picture 'crashes' the picture.) So the scenarios that we can make sense of are those associated with a non-null set of possibilities.

Moreover, since, when (relative to some background set of assumptions) if the hypothesis Δ is added to a set of hypotheses Γ , the resulting scenario cannot be made sense of, we can therefore say that (relative to those background assumptions) Γ entails $\neg\Delta$, we can also use possibilities to model entailment: Γ entails Δ (relative to certain background assumptions) just if (relative to the set of possibilities associated with those background assumptions) the set of possibilities associated with the aggregate of the hypotheses in Γ , when intersected with the set of possibilities associated with Δ , the result is the empty set of possibilities.

Finally, as we have seen, one who affirms a hypothesis is not fully coherent—can't be fully explained by Laws of rationality, in the sense of the first section—just if the hypothesis itself can't be made sense of: affirmation is in this sense 'transparent' in regard to the coherence of the hypotheses affirmed. As a result, we may use possibilities to model the attitude of affirmation as well: affirming the hypothesis that P is coherent just if the set of possibilities associated with that hypothesis is non-null; affirmation of P together with failure of affirmation of Q does not result in a fully intelligible picture of the world just if P entails Q . For this reason we may model affirmation as a relation between subjects and sets of

possibilities: to affirm the hypothesis that P is to stand in the affirmation relation to the set of exactly those possibilities at which P ; to affirm as one's total world-view the aggregate of the hypotheses in Γ is to stand in that relation to the set of possibilities intersecting the sets of possibilities associated with the hypotheses in Γ . Finally, if one fails to affirm P and fails to affirm $\neg P$, one is in this sense uncertain whether P : this is to say, in some of the possibilities in one's affirmation-set, P , and in others, $\neg P$. In this sense, uncertainty is uncertainty over which world is actual (and learning is reduction of the scope of such uncertainty).

Our notion of affirmation is *intensional*: if two subjects stand in the affirmation relation in regard to their total world-view to the set S , the subjects do not differ in any way affirmationally. For this reason our notion of affirmation may be rather more coarsely grained than the ordinary notion of *belief*: ascriptions of the latter notion are, famously, sensitive to the way in which hypotheses are packaged, whereas our notion is packaging-insensitive.

For example, perhaps *roses are red* and $(violets\ are\ blue \supset roses\ are\ red) \wedge (\neg violets\ are\ blue \supset roses\ are\ red)$ place the same conditions on the world: a scenario in which one of these is the case and the other is not the case is unintelligible. Accordingly, one affirms one of these just if one affirms the other; to affirm one but fail to affirm the other is to have a picture of the world which is not fully coherent. Of course, we can imagine it being truly said of Bill that, while he believes the former, he does not believe the latter. If we understand 'X believes that P ' along the lines suggested in the previous section as both asserting ' $TO_X(\Box P)$ ' and manifesting ' $FROM[X] : \Box P$ ', this is explicable: the matrix argument positions of 'TO' and 'FROM' are *quotational*: from a semantic perspective, these expressions operate on *discourses*, understood as linguistic objects. They do not operate on *hypotheses*; ' $\Box P$ ' appears in the logical form of 'X believes that P ' as mentioned rather than as used.

Well, that's not quite right: from the perspective of the speaker's psychology, ' $\Box P$ ' appears both as mentioned and as used. For recall that in engaging in sympathetic projection into X 's stream of consciousness—in performing $FROM[X] : \Box P$ —that stream is not completely transparent to me. I simulate 'liv-

ing through' that stream of consciousness while also treating it as an object. So if X's stream of consciousness is narrated in part by 'roses are red', when I project into it I both think that roses are red and think about the sentence 'roses are red'.

This duality is familiar in the literature on belief reports:

- Galileo believed that the earth moves. Or did he? The sentence 'the earth moves' never structured Galileo's stream of consciousness. But so what? In narrating Galileo's stream of consciousness, my primary concern is often to lay out his picture of the world, with no concern for its packaging. So I am perfectly well off in this task by performing FROM[Galileo]: the earth moves.
- Lois believes that Clark works in her office. Does she believe that Superman works in her office? Well, obviously not. The sentence 'Superman works in my office' never structures her stream of consciousness. But on the other hand, maybe so. Superman *is* Clark, so to think of Clark doing so and so is just to think of Superman doing so and so. So the hypothesis that Superman works in Lois's office is an object affirmed in Lois's stream of consciousness.

Or is it? Lois *rejects* the identity between Superman and Clark. So her space of possibilities is rather different from ours: from her point of view, there is no possibility in which Superman and Clark simultaneously occupy the same region of space; from ours, this is so in every possibility. Lois's space of possibilities is not ours plus or minus a few: every possibility Lois sees is one we regard as incoherent (in respect of its take on Clark). Short of temporarily rewiring our cognitive architecture so as to leap into an alternative modal space, we can't literally and completely interpret Lois's picture of the world 'as from within'.¹⁴

¹⁴Logic pervades the world: the limits of the world are also its limits. So we cannot say in logic, 'The world has this in it, and this, but not that'. For that would appear to presuppose that we were excluding certain possibilities, and this cannot be the case, since it would require that logic should go beyond the limits of the world; for only in that way could it view those limits from the other side as well. We cannot think what we cannot think; so what we cannot think we cannot say

The highly adventitious character of ‘TO’ and the plethora of goals we might have in engaging in FROM seem to predict the impossibility of a neat theory of the truth-conditions of belief reports, a prediction that seems to be confirmed by the intractability of debate in this literature.

These philosophical preliminaries motivate the following approach to the semantical modeling of rationality (an approach that generates what Chalmers (2005) refers to as the ‘Golden Triangle’ whereby the theory of meaning will explain why the limits of the possible and the limits of the coherent are one and the same).¹⁵

- The system begins with a number of ‘framework’ principles relating the notions of affirmation, synchronic coherence, and content understood as sets of possibilities:
 - A central psychological notion is that of *affirmation (of the hypothesis) that Φ* .
 - To model affirmation, we posit a space, \top , from which possibilities are drawn (its complement, the empty space of possibilities, is \perp). (This space may be subject-relative rather than absolute, as we saw with Lois.) One’s total affirmational state at a time can be understood as a set of possibilities $\sigma \subseteq \top$. One affirms that Φ just if $\sigma \subseteq \|\Phi\|$. In this sense, it is always the case that from one’s point of view, one’s actual possibility is a member of σ .
 - A ‘simple sentence’ ‘ P ’ is assigned a set $\|P\| \subseteq \top$, the set of possibilities in which P . Truth-functional operators can be treated along familiar lines: $\|\neg\Phi\| = \top - \|\Phi\|$, $\|\Phi \wedge \Psi\| = \|\Phi\| \cap \|\Psi\|$.

either.

‘This remark provides the key to the problem, how much truth there is in solipsism. For what the solipsist means is quite correct; only it cannot be said, but makes itself manifest. The world is my world: this is manifest in the fact that the limits of language (of that language which alone I understand) mean the limits of my world’—*Tractatus* 5.61–62.

¹⁵‘The correct explanation of the form of the proposition, ‘ A makes the judgement p ’, must show that it is impossible for a judgement to be a piece of nonsense. (Russell’s theory does not satisfy this requirement)’—*Tractatus* 5.5422.

- Coherent co-affirmation tracks coherent content: one can coherently simultaneously both affirm that Φ and affirm that Ψ just if $\|\Phi\| \cap \|\Psi\| \neq \perp$. A principle of coherent affirmation follows trivially, if we regard affirming Φ as simultaneously affirming Φ and affirming Φ : one can coherently affirm Φ just if $\|\Phi\| \neq \perp$.
- Conditional obligation to affirm tracks entailment of content: affirming Φ while failing to simultaneously affirm Ψ renders one's view incoherent just if $\|\Psi\| \subseteq \|\Phi\|$.
- A 'dynamic' approach can be useful in interpreting the indicative conditional and the certainty operator (Veltman 1996, Yalcin 2007, Starr 2010):
 - Let $\|\Phi\|_\sigma$ be the set of possibilities resulting, on pain of incoherence, from adding to one's first-person narrative the indicative sentence ' Φ ' against a prior affirmational state σ .
 - For a simple sentence ' P ', $\|P\|_\sigma = \sigma \cap \|P\|$.
 - Static and dynamic truth-functional operators behave fairly similarly: $\|\neg\Phi\|_\sigma = \sigma - \|\Phi\|_\sigma$, $\|\Phi \wedge \Psi\|_\sigma = \|\Phi\|_\sigma \cap \|\Psi\|_\sigma$.
 - The certainty operator performs a 'test' on σ to ascertain whether its matrix clause is affirmed: if so, the trivial update is performed; if not, one's affirmational state 'crashes'; as follows: if $\|\Phi\|_\sigma = \sigma$, $\|\Box\Phi\|_\sigma = \sigma$; but if $\|\Phi\|_\sigma \neq \sigma$, $\|\Box\Phi\|_\sigma = \perp$.
This provides a sense in which P and $\Box P$ are first-person equivalent.
 - The indicative conditional behaves similarly, in effect 'hypothetically testing' the result of updating σ with the antecedent, where the test performed is that of application of the certainty operator to the consequent. Let $\tau = \|\Phi\|_\sigma$. Then, if $\tau = \perp$, $\|\Phi \rightarrow \Psi\|_\sigma = \sigma$. Otherwise, if $\|\Psi\|_\tau = \tau$, $\|\Phi \rightarrow \Psi\|_\sigma = \sigma$; while if $\|\Psi\|_\tau \neq \tau$, $\|\Phi \rightarrow \Psi\|_\sigma = \perp$.
- The system could be extended to model coherence of subjective probability or 'credence' by introducing a subject- and time-indexed probability measure (a function into $[0, 1]$) over sets of possibilities:

- S affirms that P at t just if $M_S^t||P|| = 1$;
 - S has credence x that P at t — $C_S^t(P) = x$ —just if $M_S^t||P|| = x$;
 - S is coherent at t only if M_S^t is a probability measure: in particular, only if finite additivity holds, so that if P entails $\neg Q$, $M_S^t||P \vee Q|| = M_S^t||P|| + M_S^t||Q||$.
 - We could blend this with our dynamic approach by treating the familiar Bayesian rule of conditionalization as an update rule. Let $||\Phi||_\mu$ be the result, on pain of incoherence, of updating the credence function μ with adding ‘ Φ ’ to one’s first-person narrative: then $||\Phi||_\mu(\Psi) = \mu(\Psi/\Phi)$.
- Going further, we might expand the system to the modeling of explanations of action and providing a dynamic semantics for imperatives: perhaps along lines suggested in Nathan Charlow’s PhD dissertation. Considerations of space (and relevance) militate against presenting this development here.

The second part of the abstract story would concern the relation between content and packaging. In outline, it would look like this:

- We want first some concepts to allow us to talk about the packaging of content, and some framework principles to link these concepts with our apparatus for talking about content:
 - The most central notion here is that of *acceptance of a sentence*. We can think of this as a synchronic relation to a *type* of sentence. (I ignore cases where due to a certain degree of psychological fragmentation, failure to recognize the type of a certain object, or the like, one both accepts (in one partition) and fails to accept (in another partition) a certain type, or accepts a certain *token* of a certain type but rejects a certain other token.)
 - Acceptance of a sentence and affirmation of a content are related by a ‘disquotation’ schema: when ‘ P ’ is a sentence of one’s own, in ac-

cepting the sentence ‘ P ’, one affirms that P (Harman 1990, Ludlow 1999).

(This schema can really only be grasped from within the perspective of the subject of study. When I accept ‘certainly, P ’, what I affirm is very different from what you affirm in doing so. My affirmation concerns this transcendental subject, yours does not.)

Note that the conditional is not a biconditional. It had better not be! Not if differences in packaging are to do their work of explaining aspects of incoherence in one’s perspective.

Don’t think of acceptance of a sentence as affirmation of some hypothesis *about* the sentence. That need not be how acceptance and affirmation are related. Acceptance does not take place *within* affirmation; rather, affirmation takes place *through* acceptance.

I want the notion of acceptance to be somewhat strong, therefore. This complex sentence: I understand all the symbols in the sentence and the way they are put together. But is it true? I know it is either a tautology or a contradiction, but I am not sure which. I do not get the picture the sentence expresses. Now Sam comes along and tells me it is a tautology; I take her word for it, coming to believe that the sentence is true. For a trivial tautology, one the picture expressed by which I really get, Sam’s word would make no difference. This suggests that all throughout, I am thinking about the sentence, rather than thinking with it. In the intended sense of ‘accept’, I do not accept the sentence.

We want also notions of *withholding judgement* and *rejection* applying to sentences, governed by the following schemata: in rejecting ‘ P ’, one affirms that $\neg P$; in withholding judgement on ‘ P ’, one is uncertain whether P .

- We can use this disquotation schema together with relations of coherence among acts of affirmation to characterize a notion of *expression*: we may say that ‘ S ’ expresses that P just if one’s attitude toward ‘ S ’ is ‘mirrored’ in one’s attitude toward P : in accepting ‘ S ’, one thereby

affirms that P , in rejecting ‘ S ’, one thereby affirms that $\neg P$, and so forth. So for example in accepting ‘roses are red’ one thereby affirms that roses are red (by the disquotation schema); and in turn thereby affirms that if arithmetic is incomplete, roses are red; so in our sense of ‘express’, ‘roses are red’ expresses that if arithmetic is incomplete, roses are red. This is not ordinary language explication—good thing, that.

- We want finally some apparatus for talking about logical relations among sentences ‘directly’, so to speak. Our core notion here will be that of **regarding ‘ S ’ and ‘ T ’ as equivalent**.

- We want this notion to satisfy the following constraint: if ‘ S ’ and ‘ T ’ are sentences of one’s language, one regards them as equivalent just if they express the same thing.

This isn’t a reductive definition of our notion: it is rather a mutual constraint relating it to our notion of expression, and thereby in turn to our notions of affirmation, content, and coherence. The metaphysical order does not necessarily proceed from sameness of content to regard as equivalent, and neither will the order of explanation in our system always proceed in one direction rather than the other.

- Regarding ‘ S ’ and ‘ T ’ as equivalent is not just accepting ‘ $S \equiv T$ ’. Rather, it requires a more demanding pattern of attitudes toward that equivalence sentence. Following Rayo (in preparation) more or less closely, one regards ‘ S ’ and ‘ T ’ as equivalent just if one does all of the following:

1. One regards ‘ $S \equiv T$ ’ as *trivial*, as requiring nothing of the world in order to be true;
2. One regards ‘ $\neg(S \equiv T)$ ’ as *incoherent*, as not describing a comprehensible scenario;
3. One regards ‘ $S \equiv T$ ’ as *why-closed*, finding the question ‘I know that $S \equiv T$; what I want to know is *why*’ to be unintelligible;

4. One regards ' $S \equiv T$ ' (as one understands it) as true by *metaphysical necessity*: for every simple sentence ' U ' in one's language, one accepts 'even had it been the case that U , it still would have been the case that $S \equiv T$ '.
- It will not always be completely transparent to one in the moment that one regards two sentences as equivalent. To the extent that it might not be obvious that two hypotheses are the same, it might not be obvious that one regards two sentences as equivalent—not, anyway, if attitudes toward sentences are to persist through failures of coherence.

For this reason, we should distinguish regard-as-equivalent from *recognition-as-equivalent*, where the latter is something more like *active, explicit use* of sentences in entirely interchangeable ways. Regard-as-equivalent is something more like a rule or a policy, which may be followed or carried out imperfectly, but the ideal of which is recognition-as-equivalent.

Let me give an example. Suppose Bill accepts 'Hesperus = Phosphorus'. Since he thereby affirms that Hesperus = Phosphorus, he is thereby incoherent in affirming $\Phi(\text{Hesperus})$ but in failing to affirm $\Phi(\text{Phosphorus})$: after all, any two such hypotheses will be true at just the same possibilities. Suppose also that Bill accepts 'Hesperus is a planet', and thereby affirms that Hesperus is a planet. He thereby also affirms that Phosphorus is a planet, since those are the same hypotheses (even from his point of view). Since, by the disquotation rule, Bill uses 'Phosphorus is a planet' to express that Phosphorus is a planet, he thereby uses that sentence and 'Hesperus is a planet' to express the same thing. So Bill regards 'Hesperus is a planet' as equivalent to 'Phosphorus is a planet'; and, indeed, regards each simple sentence containing 'Hesperus' as equivalent to the result of substituting 'Phosphorus'.

Still, despite this, Bill is not guaranteed to recognize this. After a few beers one night, he might simply find his logic chops going to pot, and

reject 'Phosphorus is a planet'. In doing this, Bill would have gotten himself into an incoherent position, though without recognizing this. Since we grasp content through entertaining sentences, and the latter is a biological process, biological factors can result in diminished acuity in manipulating sentences, with a consequent diminution in the acuity with which we appreciate content. In such cases, the *ceteris paribus* clauses implicit in Laws of rationality are violated, and one is no longer amenable to rational explanation.

- The importance of regard as equivalent is that it is something like the 'mention' equivalent of the 'use'-like notion of *metaphysical necessity*. Let's all get together and agree to regard the sentence '⊥' as incoherent, and regard the sentence '⊤' as having a trivial update rule, OK? Then suppose Lois lets us know that she regards 'Superman = Clark' as equivalent to '⊥' and 'Superman ≠ Clark' as equivalent to '⊤'. Since we can talk in this way, we don't have to give speeches like:

According to Lois, that Superman is not Clark is metaphysically necessary. But it isn't! So evidently Lois is mistaken about the character of the proposition *that Superman is not Clark*. After all, for a proposition to be metaphysically necessary is for it to be the universal set; for it to be metaphysically impossible is for it to be the empty set. So this thing she thinks is a universal set is really an empty set. Duh! She must be confused. But how could you be confused in this way? Nothing and everything are complete opposites!! Clearly the nature of that proposition is opaque to her. She doesn't understand what she is thinking. Perhaps what is confusing her is some association she has with the sentence 'Superman is not Clark'. Maybe she associates it with the contingent proposition *that something is both a hulking guy in a cape and a nebbishy reporter*; thanks to this association, this contingent proposition blasts in front of her mind every time she considers that sentence, boggling her ability to comprehend her own picture of the world so that nothing looks like every-

thing! (Don't ask why Lois doesn't stay confused when she considers whether *that Superman is not Clark* is contingent, or whether it is a de re denial of identity.)

This sort of speech, despite its venerated position in our philosophical culture, strikes me as entirely outrageous. It seems to lose contact with the point of our use of content in laying out a person's picture of the world. It moves too quickly to the accusation of irrationality thereby undermining the connection between the theory of content and the theory of the coherent subject; requires—as is seen when the position is developed with unrelenting rigor (Chalmers 2005)—a tremendous amount of metaphysical commitment and intricate apparatus for its implementation; and threatens to leave us looking foolish if we change our mind about who Superman really is.

Rather than giving this speech, we can start out by talking about which sentences Lois regards as equivalent to one another. We can then work backward, if we manage to think ourselves into her picture through using her sentences as she uses them, to Lois's take on a subject-relative notion of metaphysical necessity. Of course we 'disquote' our own take on metaphysical necessity, advancing it as the 'real' one. But the liberally-minded should regard this take as only an embedded or 'empirical' reality: from a perspective that transcends our own as well as Lois's, our take looks more like a projection from the system of ideas we have chosen to use.

I conclude with a brief sketch of respects in which the position is or is not 'Russellian' or 'Fregean'.

- The position is Russellian in that it is *one-dimensionalist*. The content of an attitude is specified using a single set of possibilities. Moreover, these are *metaphysical possibilities*: Sam's metaphysical possibilities are the possibilities she regards as intelligible, her epistemic possibilities are the intelligible possibilities she is willing to take seriously.

- The position is Fregean in that it sees a baseline aim of a theory of content to be characterizing rationality in light of the subject's take.
- The position collapses the distinction between the Russellian and the Fregean by *making* the distinction between the 'empirically' and the 'transcendentally' real. We could think of the Fregean as departing from the position here in dissecting the empirically real into a level of reference and a level of sense, and then promoting the former to transcendental reality while the latter remains as the empirical residue. The position here regards the transcendently real as at best a matter of no significance and at worst a valorization of the empirically real for the Very Special Agent; and it attempts to suture the Fregean dissection through a rededication to the ideals of the Copernican Revolution.

3 Receptivity and justification

We are now in a position to provide a semantical theory of perceptual justification. I will develop the idea that what justifies about perception is that it is *receptive*.

Here's a rough analogy. Suppose you are at a cocktail party. As usual, you are arguing about the issues of the day with a friend—say, Lucan—and as usual, neither of you changes your mind about anything. Suddenly Paul Krugman barges in, clinking a glass as the crowd falls silent. Everyone listens intently to Krugman's long speech about the issues of the day; because of the immense trust Krugman has earned on such matters, each of you updates your picture of the world in accord with Krugman's discourse—as you understand it, and in light of your previous views. To state the obvious, Krugman's discourse stands in for the discourse of perception; your system of belief for the system of belief; Lucan's discourse for the groundless empirical discourse of the clairvoyant.

A kind of perceptual state *K* has its distinctive justificatory power, I claim, for the following reasons:

1. An instance of *K* is an *affirmation*, the content of which is packaged in a way

in which no belief can be packaged—expressed by a sentence which cannot express a belief. However, its content *can* be expressed by a sentence that is potentially accepted in a belief. When this happens, one regards those sentences as equivalent.

2. That sentence, like all sentences accepted in perceptual states and no sentences accepted in beliefs (except perhaps ‘ \top ’), is not the consequent of any Law of rationality: transition to acceptance of that sentence is always rationally inexplicable. Perception is, I shall say, *the given*.
3. The sentence with which the content of such an instance is packaged is known, as part of linguistic competence with it, to be true whenever accepted: it is (temporarily and) contingently analytic. Perception is, I shall say, *indefeasibly trustworthy*.

A state is *receptive* just if it is a *given, indefeasibly trustworthy affirmation*. What is distinctive about the rational role of perception, I claim, is that it is in this sense receptive. Clauses (1) and (2) of the characterization of receptivity are fairly anodyne, and I think we have enough in the way of apparatus to understand them without further ado. Claim (3) receives extended discussion in short order. Prior to that, however, I will show how the doctrine of receptivity explains the differences between Sam and Claire, and Sam and Fidel.

3.1 Explaining justification

If my claims are correct, then we can explain the difference in justification between Sam and Claire. By (2), Sam’s perceptual state can have no rational explanation. By contrast, Claire’s belief is of a kind which can have a rational explanation. We might want one; Claire might want one. But we won’t find it. In this sense, Claire’s belief is unjustified. To the extent that everything that can be justified should be, Claire’s belief is unjustified in a bad way (while Sam’s perceptual state is not unjustified in this bad way).

Moreover, we can explain why Sam's belief is justified but Fidel's belief is not. I argue first that if one has a belief on whether P , and one's perceptual state bears on that question, the belief should move to match the perceptual state.

By (1), if one is in K , one affirms some content P . So if one also withholds assent from, or rejects, the content P in some other state, one is incoherent. So, supposing one is in another state that takes a position on the question whether P , one is only coherent going forward if either: (a) that other state is an affirmation that P ; or (b) one abandons the state of being in K .¹⁶

Now, by (2), option (b) doesn't exist. If one is in K , that's it: being in K or not being in K is never a response to anything. So if one preserves that other state, and it isn't an affirmation that P (in accord with option (a)), one is incoherent.

One might find a strategy for living with the incoherence: if one needs to hold on to the other state and finds being in K to have suspect credentials, one might adopt a strategy of 'partitioning' or 'fragmenting' one's stream of consciousness so that the state of being in K and the other state are in different partitions and don't threaten to blow up the whole thing. But by (3), it's always a better idea *from one's own point of view* to abandon the other state in favor of, so long as one considers whether P , affirmation that P .

Sam's belief is justified, therefore, because, as the direct realist would have it, it has the same content as her perceptual state; by contrast, Fidel's belief is unjustified—and indeed should be gotten rid of—because it has a content incompatible with that of her perceptual state.

3.2 Indefeasible trustworthiness

I now turn to the question of how perception *could* be indefeasibly trustworthy: how a sentence tokened in perception could be 'temporarily analytic'.

¹⁶Quine is therefore wrong about 'recalcitrant sense-experience' being able to be incorporated without incoherence holding *any* beliefs fixed 'come what may'. Quine was doubtless thinking in terms of sense-experience as entirely inner and the beliefs in question as entirely outer. But if sense-experience is outer, even this retreat isn't available.

Consider the property Λ_ρ of *looking at the red color of a widget*. We could consider this property as a predicate type $\ulcorner \Lambda_\rho \urcorner$. We could then stipulate a referent for this predicate. In particular, we could let the referent of the predicate be the property *looking at the red color of a widget*, namely Λ_ρ itself. This predicate is therefore part of a language that is to at least this extent ‘Lagadonian’ (Lewis 1986): in a sense, the property refers to itself (although perhaps it would be more accurate to say that the property-qua-predicate refers to the predicate-qua-property). A bit more formally: $\|\ulcorner \Lambda_\rho \urcorner\| = \Lambda_\rho$.

We can extend the class of grammatical categories to include terms while preserving the Lagadonian character: consider Sam as a type of term $\ulcorner \text{Sam} \urcorner$; then stipulate that the referent of this term is Sam herself. A bit more formally: $\|\ulcorner \text{Sam} \urcorner\| = \text{Sam}$.

We could also extend the range of the two grammatical categories in the language: considering Claire (Fidel) as a type of term, let the referent of this term be Claire herself (Fidel himself); considering the property Λ_γ of *looking at the green color of a widget* as a predicate $\ulcorner \Lambda_\gamma \urcorner$, let the referent of this predicate be the property *looking at the green color of a widget*, Λ_γ .

More generally, the semantic base-clauses for our language will be the instances of the following schema: $\|\ulcorner \Xi \urcorner\| = \Xi$. The permissible substitution instances of the schema of course will be just those expressions of English (or some extension of English) which refer to entities which are semantic values of expressions in our Lagadonian language.

We can also provide a compositional semantics for the language, using a mundane story about tenseless predication of the present relative to a time t : $\|\ulcorner P(s) \urcorner\|_t = \{w : \|s\| \text{ instantiates } \|P\| \text{ in } w \text{ at } t\}$.

Finally, we will want a ‘phonetics’ and a ‘pragmatics’ for the language. Phonetics: one utters the sentence $\ulcorner \Pi(\sigma) \urcorner$ just if one = σ and one instantiates Π (permissible substituends as for the semantic based-clauses). Pragmatics: one accepts S (relative to t) just if one utters S (at t).

Let’s put this to work. Suppose that at t , Sam is looking at the red color of a widget. This is so just if Sam instantiates Λ_ρ at t . This is so just if, at t , Sam

utters the sentence $\ulcorner \Lambda_\gamma(\text{Sam}) \urcorner$ —just if Sam accepts the the sentence (relative to t) at t —just if Sam affirms at t the content relative to t of the sentence. By the compositional semantic clause, the content of this sentence is given by this set of possible worlds: $\|\ulcorner \Lambda_\gamma(\text{Sam}) \urcorner\|_t = \{w : \|\ulcorner \text{Sam} \urcorner\| \text{ instantiates } \|\ulcorner \Lambda_\gamma \urcorner\| \text{ in } w \text{ at } t\}$. By the appropriate instances of the semantic base clauses, this set is equal to $\{w : \text{Sam instantiates } \Lambda_\gamma \text{ in } w \text{ at } t\}$.

Notice that Sam’s utterance is true: in her world, Sam does instantiate Λ_γ at t . Given the way we have set up the language, we can see that this can be generalized: if one affirms a certain content through accepting a sentence of our Lagadonian language, one’s affirmation is true. We can put this by saying that Lagadonian sentences are *infallible*. And conversely: if a particular and a property are within the range of our Lagadonian language, and the particular instantiates the property, then the particular accepts the sentence of our Lagadonian language with the content that it instantiates the property, thereby affirming its content. We can put this by saying that the facts within the scope of the Lagadonian language are *self-intimating* under the Lagadonian mode of presentation. These claims sum to a *revelation* thesis: for Π and σ within the scope of our Lagadonian language, σ instantiates Π just if σ accepts a Lagadonian with the content that σ instantiates Π . More colloquially: a certain class of properties are such that one instantiates one of them just if one self-ascribes it under the Lagadonian mode of presentation.

Note that in order for one to accept a sentence S , S must be a sentence of one’s own language; and, presumably, part of what it is for a sentence to be of one’s own language is for one to have tacit knowledge of its ‘linguistic’—phonological, syntactic, semantic, and pragmatic—properties. Since the revelation thesis follows from the linguistic theory for the Lagadonian language, anyone who uses the language tacitly knows the revelation thesis.

Having explained what it would be for a language to be Lagadonian, I now advance a positive thesis, in two parts: first, there is such a Lagadonian language; and second, if σ is a subject and Π is a kind of subjective perceptual state, then σ is the semantic value of a term of that language and Π is the semantic value of a predicate of that language. For short: inarticulate sentences underlying perceptual

states are Lagadonian sentences; or: the kinds of perceptual state are in the class mentioned in the revelation thesis; or: perception is Lagadonian. I advance also a negative thesis: no (or almost no) belief involves the acceptance of a sentence of a Lagadonian language.

This then is what I mean in thesis (3), when I say that perception is indefeasibly trustworthy.

3.3 McDowell's dilemma

This explanation is only applicable in the good case, recall: we have not yet addressed hallucination, illusion, or any of the other familiar sources of difficulty for direct realism. That is our task for the remainder of the paper. However, before moving to that, I want to conclude this section by relating my story to McDowell's famous discussion of perceptual justification (McDowell 1994; my interpretation of this difficult text leans heavily on that presented in Byrne 1996).

McDowell defends the view that perceptual states have 'conceptual content'. What he means by this is not entirely clear to me, but perhaps the claim is that in a perceptual state, one affirms its content through accepting a sentence that could also underlie a belief (or perhaps one which is drawn from the language from which the sentences underlying beliefs are drawn). I worry about this position that it cannot explain the increment of justificatory power possessed by perception beyond compulsive belief, but more to the point for present purposes is McDowell's case for this position.

McDowell reaches his view by pushing back against what he regards as the view's principal historic alternatives: 'coherentism' and the 'myth' or doctrine of 'the given'. The point of this discussion is to locate our view in relation to these positions as McDowell understands them and thereby to show how it escapes the concerns he advances against them.

According to the coherentist, a system of beliefs is justified to the extent that it lacks internal sources of incoherence. McDowell worries that this view fails to respect the necessary 'discipline' imposed by the external world on a justified system of beliefs: a coherent system in the absence of such discipline would be a

mere ‘frictionless spinning in the void’.

It is evident, I hope, that our view provides no shortage of friction between the system of beliefs and the external world. The Lagadonian character of a perceptual state lashes its content to facts about the subject’s relation to the external world, and identifies its nature with these facts.

The content of both McDowell’s characterization of the doctrine of the given and his objection to it has been the subject of energetic exegetical scrutiny: what follows is my best attempt to make sense of the dialectic. Let us work backward. McDowell’s worry for the doctrine of the given is that it could explain at best why perception *causes* beliefs, and not why it *rationalizes* them: the story could provide only ‘exculpation’ when what is wanted is ‘justification’. In our framework, a story on which a theory of justification would be lacking would perhaps be a purely structural explanation of belief: one which explains how certain brain states result from certain neural processes and goes no further. Such a story could provide exculpation—beliefs might be shown to be inevitable—but lacking a first-person narrative, there would be no rational component to the story. So a story about perception which mentioned nothing about one’s affirmations would fail to be a theory of justification.

We could certainly imagine other theories that fail in this way as well: materialism and internalism are dispensable components of this failure. A story cashed solely in terms of qualia understood as objectively characterizable primitive qualities would be subject to the same worry. The classical sense-datum theory, unaided by a first-person story, on which a perceptual state is an immaterial relation between an immaterial subject and an immaterial object, would be subject to the same worry. A story cashed solely in terms of seeing understood as a causal (optical and neural) transaction between the external object and one’s ‘belief box’ would be subject to the same worry. An immaterialist externalist story with irreducible ‘edenic’ qualities of externalia in the place of sense-data would be subject to the same worry. Nor would an ‘intentional stance’ style of explanation of the sort familiar from cognitive science: the subject’s perspective is absent from such a story; the ‘representational states’ posited by such theories are mere fictions.

All of these stories abandon intentional explanation and its essential tie to the first-person, thereby surrendering on the project of explaining justification.

McDowell famously assimilates the view of Evans and Peacocke that perceptual states have ‘nonconceptual content’ to the doctrine of the given, arguing that the view and the doctrine fail for the same reason: namely, that each predicts that while one might *describe* the role of perception in the process leading to the formation of one’s belief, each also predicts that one cannot *articulate* the content of the perception. McDowell believes this because in his view thought and belief have conceptual contents: affirmations in thought and belief are underlain by acceptance of a sort of sentence which according to the competition is distinct from the kind of sentence underlying perception. This is of course a prediction of our view. So is it a problem?

Well, it is certainly the case that a view on which perception is treated as having content which is not in our sense ‘affirmed’ would face this problem. On such a view there is nothing perceptual in the stream of consciousness for belief to be a response to. We would be clairvoyants if such a story were the end of things. The mere postulation of nonconceptual content does not complete a story about justification. Conversely, postulation of conceptual content would indeed resolve the difficulty of explaining how perception could be justificatorily relevant: the sentences accepted in thought and belief underlie affirmations. But this would come at the cost discussed above.

However, there seems to be a middle way: accept a class of sentences the acceptance of which always underlies affirmation but which is distinct from the class of sentences accepted in belief. It seems to make sense to regard the class of affirmations as larger than the class of beliefs. It surely cannot be of great significance to whether a state can be rationally relevant that it is underlain by a language with a specific set of properties: after all, in transferring justification, the relation of regard-as-equivalent is indiscriminate. The commitment of the stream of consciousness to some picture or other seems to be all that is important here.

So the answer is, I think, no: the worry that our Lagadonian sentences are not ‘articulable’ in the sense that they are not thinkable is irrelevant; it would be bad

if they were not ‘articulable’ in the sense that they did not underlie affirmations; but there is no reason to suppose otherwise. If McDowell has collapsed the first-person perspective and the perspective of thought, nonconceptual affirmation of the sort found in our theory would be invisible; but *prima facie* there is no reason to collapse these perspectives. So I am inclined to think that our view is not undermined by McDowell’s attack on the doctrine of the given.

4 Into the bad

Here is our picture so far. In the good case, when one is looking at the red color of a widget, this state of one also constitutes one’s acceptance of a sentence of a certain kind ‘ \mathcal{R} ’: a kind which is true just if one is looking at the red color of a widget. This kind of sentence is not ‘articulate’, in the sense that no state of acceptance of a sentence of this kind could be a belief. If perception is to impact the remainder of one’s cognitive system, say by indirectly rationalizing an update of one’s credence function and thereby also indirectly rationalizing an alteration in one’s course of actions, perception needs to rationalize belief. Our view is that it does so via the sentential attitude of regard-as-equivalent: one regards the inarticulate sentence ‘ \mathcal{R} ’ as equivalent to the articulate sentence ‘I am looking at the red color of a widget’. Since when one regards S and T as equivalent, they express the same content, one would be incoherent in accepting ‘ \mathcal{R} ’ but failing to accept ‘I am looking at the red color of a widget’. So in order to preserve coherence in light of one’s acceptance of ‘ \mathcal{R} ’, one responds by forming the belief that one is looking at the red color of a widget.

Let us say that a belief formed as a response to a perceptual state is part of one’s *evidence*—or at least that one *treats* such a belief as evidence. And let us accordingly say that a practice by which one extracts belief from perception is an *evidential policy*. Put more briefly, then, our view is that the rational impact of perception on the remainder of the cognitive system is ‘mediated’ by evidential policy.

In the story we have told so far, evidential policy is a course of attitudes of

regarding a perceptual sentence as equivalent to a belief sentence. The story so far is too simple. Moving beyond the good case requires us to introduce complexity.

The root difficulty is that perceptual judgement can be mistaken. Sometimes, in the bad case, one is fooled: dreaming of looking at the red color of a widget, one treats the belief that one is looking at the red color of a widget as evidence. If the notion of evidence is factive, this belief does not in fact constitute part of one's evidence: it is false. Nevertheless, one treats it in the way one treats genuine evidence: say, by updating one's credential system by conditionalizing on its content. Whether such a belief deserves the unqualified appellation of the name 'evidence' is less significant, however, than that via perception, one has been led into error: one treats a false belief as evidence.

This phenomenon of 'delusive dreaming' is not compatible with the story so far. The reason is straightforward: a perceptual affirmation is always true; and two affirmations with the same content have the same truth-value; so any affirmation with the same content as a perceptual affirmation is true. But if, as on the story so far, evidential policy is exhausted by regard-as-equivalent, then any belief treated as evidence has the same content as the perceptual state to which it is a response; and so then any belief treated as evidence is true.

Our approach to this familiar issue will be motivated by reflection on the further difficulty raised by the less-discussed 'converse' of a case of delusive dreaming. In the sort of dreaming case under consideration, while one is not in fact looking at the red color of a widget, one mistakenly believes that one is. But it is also easy to envisage a 'delusive good case' in which, though one *is* in fact looking at the red color of a widget, one wrongly takes oneself to be dreaming, and therefore fails to judge that one is looking at the red color of a widget: if one is sophisticated enough, one might as a result treat a certain belief about the brain processes involved in dreaming (and not in seeing) as evidence. The difficulty here is exactly parallel: here too one treats a false belief as evidence.

We have seen three cases: the good case; the delusive dreaming case; and the delusive good case. We can round out our system of classification by recognizing cases of nondelusive dreaming: cases in which one is dreaming and recognizes

oneself to be so, and therefore does not end up with any false beliefs.

Our four types of case are generated by a pair of independent parameters. One of these concerns the kind of perceptual state one is in: *looking at the red color of a widget* versus *dreaming of looking at the red color of a widget*. The other parameter concerns something like one’s background presuppositions, perhaps as encoded in one’s prior credence distribution: whether one regards oneself as looking or rather as dreaming. Freely recombining values of these parameters generates our four cases in the way shown in this table:

		Perceptual state	
		Looking at red (\mathcal{R})	Dreaming of red (\mathcal{R}_δ)
Presupposition	Looking ^g	Good: ‘Looking at red’	Delusive dream: ‘Looking at red’
	Dreaming ^g	Delusive good: ‘?’	Nondelusive dream: ‘?’

Here the top line in each box represents the case one is in, and the bottom line represents the belief one will come to treat as evidence. The question marks in the bottom line of the bottom row represent the direct-realist friendly claim that if one takes oneself to be in the bad case, it is not immediately clear how one should regard one’s evidence as being: direct realism and the ‘transparency’ of perception often go hand in hand, and it is a plausible consequence of transparency that perceptual states do not ordinarily manifest much about their intrinsic aspects. However this issue is somewhat orthogonal to our main line of discussion, and I will set it aside. I shall use the sentence ‘?’ as a placeholder pending further discussion elsewhere of what one treats as evidence in such a case.

Notice that although what one treats as evidence varies across the cells in a *column*, it is invariant between the cells in a *row*. This invariance captures the sense that in the perceptual states under discussion are *indiscriminable* from one another (Martin 2002): subjects with the same background presuppositions will respond

cognitively to these states in the same way. At the same time, however, this indiscriminability is compatible with the possibility of correct belief or knowledge of what sort of perceptual condition one is in: in the good case, one's background presuppositions are correct, and accordingly one recognizes which kind of perceptual state one is in. Next, the contrast between the variance in a column and the invariance in a row reflects the Cartesian sense that empirical error never results from perception, but rather only ever from a failure to treat perception in a way appropriate to extracting the truths contained within it. Finally, although the table does not depict this fact, the dependence of evidence on presuppositions is compatible with a further dependence on the perceptual state: two subjects may share the presupposition that they are looking rather than dreaming but end up with distinct evidence if the perceptual state of one is a case of looking at *red* while that of the other is a case of looking at *green*.

The question we face, though, is how to explain the table: what the 'rational dynamics' involved in the generation of empirical belief, captured in evidential policies, are such that our pattern of reactions is as depicted in the table.

My suggestion will be that an evidential policy is *conditional*: not in force absolutely, but only relative to the truth of a certain hypothesis. To see what I mean here consider a contrasting case of a pair of 'logical policies', policies of regarding belief sentences as equivalent. If one accepts DeMorgan's law, then for any sentence of form ' $P \wedge Q$ ', one's policy is to treat it as equivalent to ' $\neg(\neg P \vee \neg Q)$ '. This policy is *absolute*: one to be followed no matter what other conditions may hold. By contrast, one might introduce a certain modicum of externalism to one's understanding of a term by relativizing one's logical policies in regard to it to conditions among which one may be uncertain. 'Water', they say, is a 'natural kind term': ordinarily it is used to refer to whatever the underlying chemical kind may be that manifests in the watery way. If this is true, it could be modeled by attributing to its users a class of conditional logical policies:

- Regard ' $\Phi(\text{water})$ ' and ' $\Phi(\text{H}_2\text{O})$ ' as equivalent *if* H_2O manifests in the watery way; and
- Regard ' $\Phi(\text{water})$ ' and ' $\Phi(\text{XYZ})$ ' as equivalent *if* XYZ manifests in the

watery way; and

- ...

Or, put as a schema:

- Regard ‘ $\Phi(\text{water})$ ’ and ‘ $\Phi(\Pi)$ ’ as equivalent *if* Π manifests in the watery way,

where the permissible substitution-instances of ‘ Π ’ are chemical kind terms.

If one has this conditional policy, then one accrues by one’s lights a certain class of obligations when one accepts ‘water is tasty’: for example one binds oneself to accepting ‘XYZ is tasty’ if XYZ manifests in the watery way. This sort of binding is such that what exactly is required in order to comply with it may not be obvious to one so bound. As a matter of fact H_2O manifests in the watery way, so if one strikes the conditional policy and accepts ‘water is tasty’, one is bound to accept ‘ H_2O is tasty’ and thereby incoherent by one’s own lights in failing to do so. This sort of incoherence is of course explicable by one’s ignorance of what is required for carrying out one’s own policy, and so does not falling into it is not exactly as bad as going stark gibbering mad. However, it does reflect a certain disequilibrium in one’s cognitive position, which can be remedied by learning some chemistry and enlarging the class of sentences one accepts in the appropriate manner.

Suppose that one has the following conditional evidential policies:

- Regard ‘ \mathcal{R} ’ and ‘I am looking at the red color of a widget’ as equivalent *if* I am looking;
- Regard ‘ \mathcal{R}_δ ’ and ‘?’ as equivalent *if* I am dreaming.

These policies predict one’s reactions in the ‘on-diagonal’ cells: the cells in which one’s presuppositions are accurate. In the good case, policy (A) not only applies but is ‘triggered’ by one’s belief that one is looking: in that case, one regards *I am*

looking at the red color of a widget as evidence, as desired. And in the nondelusive dream case, policy (B) is triggered by one's belief that one is dreaming: in that case, one regards some different hypothesis—whatever it is—as evidence, as desired.

This conditionalized story of the dynamics of perceptual evidence improves on the earlier story. The difficulty for that story was its prediction that in all cases, when one comes to treat a hypothesis as evidence, this is the result of a policy of treating the sentence expressing it as equivalent to one's perceptual sentence. Since one's perceptual sentence is always true, this had the consequence that one only ever treats a truth as evidence. And in the 'off-diagonal' cells, where one's presuppositions about one's perceptual state do not align with the facts, that consequence is not correct.

The improvement over the earlier story comes about because in the off-diagonal cells, neither policy (A) nor policy (B) issues any requirement. Suppose one is in the delusive dream case: then policy (A) is triggered by one's belief that one is looking, but because one accepts in perception not ' \mathcal{R} ' but ' \mathcal{R}_δ ', the policy is silent in regard to what belief sentence one must come to accept. Conversely, while policy (B) would if triggered require one in light of one's acceptance of ' \mathcal{R}_δ ' to accept '?', policy (B) is not triggered.

So far so good: a false prediction has been eliminated. But it has not yet been replaced with a true prediction. The conditionalized theory as developed so far makes no prediction about what will happen in the off-diagonal cells. The story requires further supplementation.

I begin by describing an approach I think the direct realist should reject. Consider these more complex conditional policies:

- a. (i) Regard ' \mathcal{R} ' and 'I am looking at the red color of a widget' as equivalent *if* I am looking;
- (ii) Regard ' \mathcal{R} ' and '?' as equivalent *if* I am dreaming;
- b. (i) Regard ' \mathcal{R}_δ ' and '?' as equivalent *if* I am dreaming;

- (ii) Regard ' \mathcal{R}_δ ' and 'I am looking at the red color of a widget' as equivalent *if* I am looking.

Two facts about these policies are direct-realist unfriendly. First, ' \mathcal{R} ' and ' \mathcal{R}_δ ' are not distinguished: clauses (a.i) and (b.ii) treat the sentences the same way if one is looking, while clauses (a.ii) and (b.i) treat them the same way if one is dreaming. In this sense they are equivalent in meaning, so that the character of one's perspective is the same whichever of them is tokened. But in the view of the direct realist, whether one is connected perceptually to one's environment influences the character of one's perspective. Second, the metaphysics of perceptual states takes an odd turn under these policies. Since the belief one accepts as evidence is the truth about one's perceptual state, and ' \mathcal{R} ' and ' \mathcal{R}_δ ' are in effect kinds of perceptual state, it follows that each of these is a sort of odd kind of relative property: when instantiated when looking, ' \mathcal{R} ' is a certain kind of looking, but when instantiated with dreaming, ' \mathcal{R} ' is something else entirely. The direct realist might be happy that in the good case one's perceptual state includes the external world, but accepting that in a certain sense that very same state is also presentable in the bad case is less likely to go down. In light of the bizarre character of the metaphysics involved in the proposal, the direct realist is not the only one that should eschew this approach.

The failure of the complex conditionalized approach is of more than merely technical interest. After all, it seems to be the only way of using our apparatus to predict that in each of the four cells, one acquires one's evidence by correctly following one's evidential policies. So we must conclude that in delusive cases, one's evidence is the product of a *mistake* about what one's policies require of one.

Here is one way of implementing this idea. Suppose that one's evidential policy (in regard to the cases under consideration) is exhausted by the conditional policies (A) and (B). One's evidential policy therefore contains no clause of either of the following forms:

- Regard ' \mathcal{R} ' and 'blah blah blah' as equivalent *if* I am dreaming;

- Regard ' \mathcal{R}_δ ' and 'blah blah blah' as equivalent *if* I am looking.

The direct realist should in fact be happy with this result: since one cannot even token ' \mathcal{R} ' if one is dreaming, a policy stipulating what to do with it when one is dreaming would be in a certain sense defective. What to do if one tokens ' \mathcal{R}_δ ' while under the impression one is looking? What should one's policy be here? The question is not easily posed from the first-person perspective. If one is under the impression one is looking, then from the first-person perspective things are this way: I am looking. Fixing this, the question of what to do if one tokens ' \mathcal{R}_δ ' is then a question of what to do in an incoherent situation. Rationalizing policies and rules provide answers about what to do if things are this or that way; given a way things can't be, such policies are silent. (Alternatively we could implement this idea using an unconditional evidential policy which, like all policies, only gives advice of how to respond to the world when the world is coherent.)

At this point, we see what I take to be the root of philosophical perplexity about perception. In a delusive case, *one's perspective is incoherent*: the perceptual aspects of one's perspective affirm a certain hypothesis; the doxastic aspects affirm a certain incompatible hypothesis. In such circumstances, all bets are off from the point of view of intentional psychology: *ceteris* ceases to be *paribus*; explanatory factors must shift away from aspects of one's perspective to more base matters of fact.

Such an explanation might have a form something like the following. Let us say that one *habitually reacts alike* to a certain class P of perceptual sentences just if, as a matter of habit—taken to include but not be limited to cases where this is a matter of policy—throughout a range of cases broader than those within which intentional psychology is explanatorily efficacious, throughout a wide class B of belief sentences one might accept, fixing acceptance of one of the B but varying which of the P one accepts against the background of acceptance of that B does not vary which belief sentence one comes to accept as a matter of 'primitive compulsion'. For example, one habitually reacts alike to ' \mathcal{R} ' and ' \mathcal{R}_δ ' in the sense that, as a matter of habit, even in circumstances in which one's perspective is incoherent: if one accepts 'I am looking', one's immediate doxastic reaction is

the same whether one accepts ‘ \mathcal{R} ’ or ‘ \mathcal{R}_δ ’; and the same is so if one accepts ‘I am dreaming’.

We can use this definition in wrapping up our explanation of the patterns in our matrix. Why do we observe the phenomena we do in delusive cases? We answer as follows: because one habitually treats ‘ \mathcal{R} ’ and ‘ \mathcal{R}_δ ’ alike; and then we stop speaking.

Now, this answer rightly appears to have a certain sort of ‘*virtus dormativa*’ character. To label a class of perceptual sentences as one to the members of which one habitually reacts alike is not of course to provide any deeper explanation, say in neural terms, of this habit. And yet, the story can be seen to be a very healthy one for other reasons. Labeling a class of perceptual sentences in this way does immediately display the necessity of launching project of providing non-psychological explanations of what it is about such sentences that makes one habitually react alike to them: in this sense, our definition is ‘fruitful’. Perhaps more importantly, the explicit treatment of cases outside the boundary drawn around the psychological by the definition shows the possibility of limiting our ambitions and giving different explanatory domains their due: sometimes when one habitually reacts alike to a class of sentences, there is just no intentional explanation of this fact in the offing; the sense that there must be should no longer corrupt our investigations into just what it is that intentional explanation looks like. The ‘*metaphysics of perception*’, if understood as a project of providing psychological explanations of common doxastic reactions to seeing and hallucination, is dissolved as a project.

Let me sum up. We provide a two-tiered story of evidential dynamics. In the privileged core, one’s presuppositions about one’s situation do not clash with the contents of one’s perceptual states. Under such circumstances, one’s evidential policies provide one with a correct interpretation in the language of belief of what is accepted in perception. One treats as evidence a sentence that is, indeed, a piece of evidence.

Outside of the privileged core, one’s presuppositions do clash with one’s perceptual states, rendering one’s perspective incoherent. Under such circumstances,

evidential policies no longer have any force. If one treats a certain sentence as evidence in light of one's presuppositions and perceptual states, there is no intentional explanation of this fact: no more than there is an intentional explanation of Bill's drunken failures of logic. One's reactions in a delusive case are not explicable by Law.

This paper concerns the relation between direct realism and perceptual justification, so I should make explicit the claims about justification implicit in this two-tiered story. In the privileged core, one's perceptual state justifies the beliefs one treats as evidence: in light of one's evidential policy, one would be incoherent not to affirm this belief if in that perceptual state. So we can recognize what it is that one sees in forming the belief: being in the perceptual state requires it. In that sense, one is justified in forming the belief.

By contrast, outside the privileged core, considerations of rationality do not apply. In this sense, the belief one treats as evidence is not *justified* in the delusive good case or the delusive dream case. At the same time, however, it is certainly *natural* in the delusive dream case to believe that one is looking at the red color of a widget. If I were in that case, I would—as it happens—do the same (just as if I were as drunk as Bill, I would make gross failures in logic). So in that more attenuated sense, one's belief is 'justified' even here. (Note that the position is therefore a version of 'evidential externalism' in the sense of Silins 2005: the good case provides a higher increment of justification to one's belief than the delusive dream case. But the increment is not, as Silins suggests, an increase in subjective probability: rather, it is the presence of both rational and natural justification, rather than merely the latter.)

In a sense, then, my story about delusive cases offers exculpation rather than justification. But, to turn McDowell's famous phrase on its head, it would be a mistake to attempt to offer justification: exculpation is just what is wanted.

5 Extensions

5.1 Illusion

Just as hallucination both can be delusive and can be nondelusive, so-called ‘illusory’ perception both can be delusive and can be nondelusive.

Here is a case of ‘illusory’ perception: suppose a white widget is rolling by Ilya, at each moment carefully illuminated by a red spotlight. In looking at the color of the widget, Ilya accepts a certain perceptual sentence ‘ \mathcal{R}_1 ’. Suppose this sentence is one we habitually treat alike with ‘ \mathcal{R} ’ and ‘ \mathcal{R}_δ ’. What Ilya comes to treat as evidence depends on his background assumptions: accepting either that he is dreaming or that things are in no way tricky would result in the exculpated but unrationalized beliefs we have seen already. But suppose that Ilya adds to the conditional policies (A) and (B) also the following:

- C. Regard ‘ \mathcal{R}_1 ’ as equivalent to ‘I am looking at the red illumination and white color of a widget’ *if* I am in a situation with the odd features that such-and-such;

in that case, if Ilya presupposes he is in a situation with the odd features that such-and-such, he will come to rationally regard as evidence that he is looking at the red illumination and white color of a widget.

On my story, there is a sense in which veridical perception sets the standard for the cases that of been of central interest to philosophers. If we know that certain sorts of seeing, hallucination, and illusion are habitually treated alike, and we know how one rationally reacts to that sort of seeing given appropriate background beliefs, then we can piggyback on this understanding our predictions about how one with those same background beliefs will react to those sorts of hallucination and illusion. Such a story would of course be only an ‘as if’ story rather than actually reflecting what goes on with one. Despite this, it might be instrumentally of value in predicting reactions.

5.2 The de re

Tweedledee and Tweedledum are identical in height, weight, body type, hair color and style, affect, grace, and countenance: for the most part, what can be seen of one in his intrinsic aspects can be seen of the other. As a result, they are easily mistaken for one another.

Tina knows Tweedledee very well but has never met Tweedledum, and indeed is unaware of his existence. Whenever she looks at someone's 'tweedlish' aspect, then, she treats as evidence 'I am looking at Tweedledee'. We may then say that Tina's evidential policies include a wide variety of instances of the following schema:

- D. Regard ' T ' as equivalent to ' $f(T)$ ' if I am in the good case and no one exactly resembling anyone I know is around;

Where permissible substitution instances of ' T ' are perceptual sentences in which one is looking at Tweedledee's tweedlish aspect, and ' $f(T)$ ' is a belief sentence entailing 'I am looking at Tweedledee'.

Now of course in light of the fact that Tweedledee and Tweedledum are twins, we habitually treat perceptual sentences concerning them alike; accordingly, if Tina were to see Tweedledum, she would affirm as evidence that she is looking at Tweedledee. According to our theory, this affirmation is exculpable but not rationalizable.

5.3 Seeing-as

Young Tycho looked to the heavens and learned little or nothing of interest; Old Tycho looked to the heavens and instantly spotted a great range of detail about where in an intricate pattern of heavenly movement things presently were located. Before enrolling in a botany course, Alfred couldn't tell a pine from an oak and didn't even try; afterward he was able to identify pines just by looking.

Perhaps then the change in Alfred amounts to his coming to adopt a certain evidential policy: namely, where ' \mathcal{P} ' is a perceptual sentence with the content that one is looking at a pine,

E. Regard ' \mathcal{P} ' as equivalent to 'I am looking at a pine' *if* I am in the good case and nothing botanically weird is going on.

Byrne (forthcoming) raises the following worry for a closely related view developed by Siegel (2006):

Imagine that pines grown on Island A look like normal pines, and that pines grown on Island B look like oaks (due to the strange soil and climate). One develops a recognitional disposition for the tree on Island A, and similarly for the tree on Island B (but does not know that the trees are identical). If Siegel's argument works, then if one sees an A tree and a B tree side by side, they will both be visually represented as pines. Presumably, then, (a) one will believe that the two trees are of the same kind, and (b) they will appear *more* visually similar after one has learned to recognize them by sight than they did before. Clearly neither of these predicted consequences will be borne out.

Let's say that Alfred only tokens ' \mathcal{P} ' when looking at an A-pine; looking at a B-pine, he tokens ' \mathcal{P}_β ' instead.

In order to get a prediction here we need a story about Alfred's evidential policy in regard to ' \mathcal{P}_β '. One such story is a story of delusion: from his botany course, Alfred has the following evidential policy in regard to a certain perceptual sentence ' \mathcal{O} ':

F. Regard ' \mathcal{O} ' as equivalent to 'I am looking at an oak' *if* I am in the good case and nothing botanically weird is going on;

since something weird is going on, Alfred's position is incoherent. We get a prediction about his beliefs if we assume that Alfred habitually treats ' \mathcal{O} ' and ' \mathcal{P}_β ' alike; we then predict he is exculpated in falsely thinking a B-pine is an oak.

Another story about Alfred's evidential policy restores us to the nondelusive. This 'direct' evidential policy governs ' \mathcal{P}_β ':

G. Regard ' \mathcal{P}_β ' as equivalent to 'I am looking at a bine' *if* I am in the good case and nothing botanically weird is going on.

Then, when Alfred looks at a B-pine, he justifiably accepts 'I am looking at a bine'. What does this mean? Well, 'bine' is not a term in our vocabulary: it is Alfred's term for the B-pine, which we know to be a sort of 'race' of pine, but from Alfred's point of view is a species of tree about which he is *de re* uncertain whether it is identical to the pine.

Neither of these proposals makes either of Byrne's predictions. Perhaps Byrne is presupposing that the structure of the world as Byrne understands it directly injects itself into the structure of the world as Alfred understands it: that Byrne's certainty that the bines are the pines is also mirrored in the certainty of anyone in a position to think about bines and pines—or at least that any Millian must think so (Byrne follows up this discussion by considering a Fregean rebuttal). This way of thinking about the Millian apparatus strikes me as profoundly misguided: the point of intentional characterization and explanation is to understand the world from the subject's point of view, rather than from the view of some hybrid of the subject and oneself or total science or whatever.

A final question about this proposal is whether Alfred was in a position to entertain ' \mathcal{P} ' prior to his botany class. Perhaps so: perhaps in all 'perceptual' respects Alfred is the same looking at a pine tree before and after, but in the past he did not discriminate in his treatment of ' \mathcal{P} ' and the perceptual sentences he accepted when looking at other tree-like objects; in this case, we would say that Alfred's perceptual sentence ' \mathcal{P} ' underwent a subtle shift in meaning, along the lines of the shift sentences involving theoretical terms undergo when the stipulative terms of the theory undergo subtle modification. Perhaps not: perhaps taking the botany course resulted in a change in Alfred in 'perceptual' terms, so that now he is but before he was not able to entertain ' \mathcal{P} '.

But perhaps there is no real difference between these proposals: our apparatus of perceptual sentences and evidential policies is a theoretical tool useful in various ways, but we should not get too hung up in the assumption that all questions that can be posed using the apparatus are ones that deserve answers (*viz.* Benac-

erref 1965). This is not to deny of course that questions about what happens at the ‘implementational’ level in cases of seeing-as are interesting or answerable: rather that intentional psychology may be silent here.

5.4 Uncertain evidence

Val is uncertain whether he is awake or dreaming. Suppose that as a matter of fact he is awake, and accepts ‘ \mathcal{R} ’. Suppose moreover that Val’s evidential policies include (A) and (B): recall,

- A. Regard ‘ \mathcal{R} ’ and ‘I am looking at the red color of a widget’ as equivalent *if* I am looking;
- B. Regard ‘ \mathcal{R}_δ ’ and ‘?’ as equivalent *if* I am dreaming.

What will happen with Val?

Well, Val is uncertain whether he is awake or dreaming. Policy (A) triggers only if Val is awake; policy (B) triggers only if Val is dreaming. Now, note that the triggers for these policies are *objective* conditions rather than subjective conditions: the trigger for (A) is not one’s *believing* oneself to be awake, but one’s being awake. This is important for Val, because the subjective condition would not trigger his policy at all: he does not believe himself to be looking. Rather, the trigger is that one *is* looking. So Val is uncertain whether his policy has triggered. What should one do when one is uncertain which of one’s policies has triggered?

One notional option is to regard this as a garden variety case of decision-making under uncertainty: weigh up the expected costs and benefits of the various options, perform the one with the highest expected value, and hope for the best. But the case here is poorly suited to assimilation to this approach, in two ways. First, this is *evidential* policy we are concerned with: policy about that on which to conditionalize one’s prior credence function. Conditionalizing on one or the other judgement—looking at the redness of a widget versus ?—drives out the sort of uncertainty that generated the decision in the first place. Ordinary decision making under uncertainty is not like this: in choosing for the best, one remains

aware of one's possible failure. Bayesians should be wary of recommending approaches that generate certainties as voluntary choices from states of uncertainty for a number of reasons, not the least of which is that evidential uncertainty would be a peculiarly unstable sort of state. And second, this is a *cognitive* policy we are talking about here. Prima facie, there is no reason a single decision must be made here. Eventually, one must come down on exactly one color of wine (given one's budgetary constraints), but there is plenty of room in logical space for pictures of the world.

It strikes me therefore that a better approach for Val would be to be indeterminate in which evidential policy he carries out. He should be in a state in which he somehow carries out (A) 'to degree .6' and somehow carries out (B) 'to degree .4'. What would this look like?

Suppose Val's initial credence distribution is the function C , and that C conditionalized on *I am looking at the red color of a widget* (in accord with policy (A)) is C_A , while C conditionalized on ? (in accord with policy (B)) is C_B . Val should then go in a state in which he 'commits' to degree .6 to C_A and to degree .4 to C_B . Just to have a helpful label, let us make the following distinction. *Uncertainty* is the notion captured in the traditional account of credence distributions: when any of a range of worlds might be actual, one is uncertain whether P just when at just some of the worlds that might be actual, P . Our novel notion could be called *unclarity*: partial, perhaps graded, commitment to a range of mutually incompatible states of uncertainty.

This 'partial commitment' approach is unlike the approaches to evidential uncertainty found elsewhere in the literature. The traditional approach, involving 'Jeffrey Conditionalization', involves coming to ignore the worlds outside of the disjunction of *I am looking at the red color of a widget* and ?, and somehow redistributing probabilities among the remainder of the worlds such that 60% of probability is over the former. The difficulty here, pressed by Williamson (2000, 216–7), is that the 'somehow' does all the work: we are not given a formal approach to updating, but rather a loose parametrization of such an approach.

Williamson's alternative (198–9) is in effect to eliminate evidential uncertainty, finding some hypothesis about which one is certain as one's genuine evidence: perhaps the hypothesis that it looks to one as if one is looking at the red color of a widget. This approach faces three concerns, however. The first of these is internal to our project of developing a theory of evidence for the methodological direct realist. Direct realists would be uncomfortable with the eventual availability of such an internal hypothesis in all cases: why if so would it not do all the work in all cases, perhaps abetted by one's prior credences? The second concern is more broadly distressing, albeit programmatic. In my view, claims like 'it looks as if P ' and ' o looks F ' have a logical form something like ' $[\exists Q : \text{if } P, Q](\text{look}[Q])$ ' or ' $\text{look}[o \text{ is } F]$ '. The meaning of the 'look' operator, in turn, is to perform a 'test' on whether its complement is entailed by what one regards as one's evidence (rather: a certain subset of that, namely the subset arrived at by looking). This meaning appeals to what one regards as evidence. So appeal to a notion like how things look in characterizing the nature of one's evidence must be circular.

The third concern is that for a range of other reasons we want partial commitment or unclarity, at some of which we have already hinted in this paper. First, one might be uncertain about the 'objective chance' of a certain proposition: perhaps by being fairly confident that this coin is fair, somewhat doubtful that it is biased to degree .75 toward heads, and barely willing to take seriously that it is biased to degree .6 toward tails, and therefore uncertain about the chance associated with the hypothesis H : *the next time this coin is flipped it comes up tails*. A simple chance-credence principle would say something like 'if one believes that the chance of P is x , one should have credence x that P '. If one were certain of fairness, the simple principle would recommend credence .5 in H ; if certain of .75 heads bias, credence .75 in H ; if certain of .6 tails bias, credence .4 in H . So applying the simple chance-credence principle to one's uncertain state would suggest 'strong confidence' in credence .5, 'some doubt' in credence .75, and 'bare recognition' of credence .4. *Perhaps* what this mixed state amounts to is simply a credential state that aggregates these several credential states, by taking a weighted average of their assignments (Lewis 1980): this would reduce the sort

of unclarity in this case to simple uncertainty. Still, perhaps not: treating unclarity about chance as epistemic possibility of a range of different assignments to a certain magnitude precludes an expressivist treatment of statements of objective chance, thereby rendering chance-credence principles difficult to justify a priori. In any event, perhaps such heavy machinery is not required to get the desired result: perhaps simple vagueness of subjective probability is best treated as partial commitment.

However, other sources of unclarity are not so easily finessed away. For instance, one might be uncertain whether Hesperus is Venus or Mars. As we understand this, that is a matter of uncertainty whether to regard as equivalent sentences of form ' $\Phi(\text{Hesperus})$ ' and ' $\Phi(\text{Venus})$ ' or rather to regard as equivalent sentences of form ' $\Phi(\text{Hesperus})$ ' and ' $\Phi(\text{Mars})$ '—where one regards as *inequivalent* such sentences as 'Venus is closer to the Sun than Earth' and 'Mars is closer to the Sun than Earth'. Since when regards a pair of sentences as equivalent, they have the same content—are necessarily coextensive from one's point of view—there can be no single credence distribution involving worlds in which Hesperus is Venus and also worlds in which Hesperus is Mars.

Alternatively, one might be uncertain whether intuitionism, classical logic, or dialetheism is correct. Credence distributions are probability functions, and a probability function assigns every logical truth the value 1. So uncertainty about which sentences of one's own—of which one knows the meaning—are the logical truths can therefore manifest in uncertainty about whether to assign credence 1 to the hypothesis expressed by a certain sentence. This cannot be 'uncertainty' of the standard sort but must rather be something like our unclarity: partial commitment to each of a range of credence functions.

These three sorts of cases have the following in common with ours: each case involves indecision about not which world one is in, but about subjective probability itself. In the case of indecision about chance, we regard the chance-credence principle as falling out of an account of chance sentences as expressions of subjective probability, so that indecision about chance is indeterminate subjective probability. In the case of indecision about identity, we regard identity sentences

as expressing patterns of epistemic necessity, so that indecision about such a sentence is indeterminacy in whether various propositions are epistemically necessarily coextensive. In the case of indecision about logic, we regard logical truths as subjectively certain, so that this sort of indecision is unclarity about the scope of the epistemically possible. Finally, our treatment of evidential uncertainty as unclarity sees indecision about what to presuppose in regard to one's epistemic position as ramifying to indecision about what one's evidence is; since one is certain of what one regards as evidence, this sort of indecision is also unclarity about the scope of the epistemically possible.

(What are the objects of credence functions on the sort of partial commitment approach under consideration? From the outside, they look like sentences; from the inside, like hypotheses.)

But of course the question remains: how shall we understand this notion of partial commitment to a credence function? Providing a preliminary functional specification would require two articles of mathematics: a confirmation theory and a decision theory. The former may be a relatively trivial extension of the standard Bayesian apparatus. But I do not expect the latter to be trivial at all. Either way, the necessary development will have to wait for a different forum.

References

- Ayer, Alfred J., editor, 1959. *Logical Positivism*. New York: The Free Press.
- Benacerraf, Paul, 1965. 'What Numbers Could Not Be'. *The Philosophical Review*, 74:47–73.
- BonJour, Laurence, 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Byrne, Alex, 1996. 'Spin Control'. In Enrique Villanueva, editor, *Perception*, volume 7 of *Philosophical Issues*, 261–74. Atascadero: Ridgeview.
- Byrne, Alex, forthcoming. 'Experience and Content'. To appear in *Philosophical Quarterly*.
- Carnap, Rudolf, 1932. 'Psychology in Physical Language'. *Erkenntnis*, 3:***–***. Reprinted in Ayer 1959.
- Chalmers, David J., 2004. 'Imagination, Indexicality, and Intensions'. *Philosophy and Phenomenological Research*, 68:182–90.
- Chalmers, David J., 2005. 'The Foundations of Two-Dimensional Semantics'. In Manuel García-Carpintero and Josep Macià, editors, *Two-Dimensional Semantics: Foundations and Applications*. Oxford: Oxford University Press.
- Gillies, Anthony S., 2001. 'A New Solution to Moore's Paradox'. *Philosophical Studies*, 105:237–50.
- Graham, Peter A., 2010. 'In Defense of Objectivism About Moral Obligation'. *Ethics*, ***:***–***.
- Harman, Gilbert, 1990. 'Immanent and Transcendent Approaches to the Theory of Meaning'. In Roger Gibson and Robert B. Barrett, editors, *Perspectives on Quine*. Oxford: Blackwell. Reprinted as "Immanent and Transcendent Approaches to Meaning and Mind" in Harman 1999.
- Harman, Gilbert, 1999. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.
- Heal, Jane, 2003. *Mind, Reason, and Imagination*. Cambridge: Cambridge University Press.

- Hellie, Benj, 2010. 'An Externalist's Guide to Inner Experience'. In Bence Nanay, editor, *Perceiving the World*. Oxford: Oxford University Press.
- Hellie, Benj, in preparation. 'Conscious Life'. MS, University of Toronto; forthcoming with OUP.
- Lewis, David, 1980. 'A Subjectivist's Guide to Objective Chance'. In Richard C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume II. Berkeley: University of California Press.
- Lewis, David, 1986. *On the Plurality of Worlds*. London: Blackwell.
- Ludlow, Peter, 1999. *Semantics, Tense, and Time: An Essay in the Metaphysics of Natural Language*. Cambridge, MA: The MIT Press.
- Martin, Michael G. F., 2002. 'The Transparency of Experience'. *Mind and Language*, 17:376–425.
- McDowell, John, 1994. *Mind and World*. Cambridge, MA: Harvard University Press.
- Rayo, Agustín, in preparation. 'Possibility and Content'. MS, MIT.
- Ross, W. D., 2002. *The Right and the Good*. Oxford: Oxford University Press.
- Siegel, Susanna, 2006. 'Which Properties are Represented in Perception?' In Tamar Szabó Gendler and John Hawthorne, editors, *Perceptual Experience*. Oxford: Oxford University Press.
- Silins, Nicholas, 2005. 'Deception and Evidence'. *Philosophical Perspectives*, 19:375–404.
- Starr, William, 2010. *Questions and Conditionals*. Ph.D. thesis, Rutgers University, New Brunswick, NJ.
- Thompson, Evan, 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.
- Thompson, Michael, 2008. 'Naive Action Theory'. In *Life and Action*. Cambridge, MA: Harvard University Press.
- Varela, Francisco, 1991. 'Organism: A Meshwork of Selfless Selves'. In A. Tauber, editor, *Organism and the Origin of Self*. Dordrecht: Kluwer.

- Veltman, Frank, 1996. 'Defaults in Update Semantics'. *Journal of Philosophical Logic*, 25:221–61.
- Williamson, Timothy, 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig, 1921/1974. *Tractatus Logico-Philosophicus*. Atlantic Highlands, NJ: Humanities Press. Translated by David F. Pears and Brian F. McGuinness.
- Yalcin, Seth, 2007. 'Epistemic Modals'. *Mind*, 116:983–1026.