

Epistemological Consequences of Frege Puzzles

Timothy Williamson

Models in formal epistemology—such as models of epistemic logic and Bayesian probabilistic models—tend to be *coarse-grained*: they obliterate cognitively significant differences for the sake of mathematical simplicity and power. By contrast, ascriptions of knowledge and belief in natural language look more *fine-grained*: they seem sensitive to those very cognitive differences. However, developments in the semantics of natural language threaten to undermine that contrast, by analysing the truth-conditions of propositional attitude ascriptions as less sensitive to cognitive differences than they seem. These issues matter for epistemology, because it is mostly done in natural language: even formal epistemologists use natural language to explain the intended applications of their mathematical models. If discourse in natural language about knowledge, belief, and other epistemologically interesting relations does not work quite as it seems to, then arguments in epistemology may be led astray by misleading appearances. For instance, when testing an epistemological generalization involves assessing a sentence of the form ‘S knows that p’ as true or false in a hypothetical case, our assessment may go wrong because we confuse semantic and pragmatic aspects of knowledge ascriptions in English. Although indiscriminate scepticism about our assessments would be unwarranted, they may need some fine-tuning.

This article explains and explores the issue in detail, though far from comprehensively. As will emerge, there are specific reasons to postulate certain kinds of error in our attitude ascriptions.¹ The article’s conclusions will be tentative. Its main purpose is consciousness-raising: if we want to do epistemology accurately, we must take proper account of these issues.

1. *The Fregean consensus*

In the old days, philosophers took it as a datum that someone can believe that Hesperus is Hesperus without believing that Hesperus is Phosphorus. Thus substituting co-referential proper names in the ‘that’-clause of a belief ascription is not always truth-preserving. Presumably, one who believes the obvious truth that Hesperus is Hesperus also *knows* that Hesperus is Hesperus, while one who fails to believe that Hesperus is Phosphorus also fails to *know* that Hesperus is Phosphorus, because knowing requires believing. Hence someone can also know that Hesperus is Hesperus without knowing that Hesperus is Phosphorus. Thus substituting co-referential names in the ‘that’-clause of a knowledge ascription is also not always truth-preserving. Ascriptions of other propositional attitudes such as wondering, doubting, hoping, and fearing behave likewise.

Some version of Frege's distinction between sense and reference was widely (though not universally) taken to explain this phenomenon. The names 'Hesperus' and 'Phosphorus' have different senses, different modes of presenting the same planet, which their occurrences contribute as components of the senses of simple sentences in which the names occur; the senses of the sentences are Fregean thoughts or propositions. Thus the sentences 'Hesperus is Hesperus' and 'Hesperus is Phosphorus' express different propositions. There is no obstacle in principle to having an attitude to one proposition without having it to the other. On Frege's own version of the view, words in the 'that'-clause of an attitude ascription refer to their usual *senses*, not their usual *referents*, so the names were not even co-referential in the context in which the substitution was made, though the general consensus did not extend to that reference-shifting mechanism. The consensus was just that some account or other of the semantics of attitude ascriptions would explain how the propositional content of the ascribed attitude depends on the ordinary senses, not the ordinary referents, of expressions in the 'that'-clause.²

Given that consensus, Frege puzzles presented no special danger to epistemology. One had to be careful not to make illicit substitutions when characterizing what was putatively known or believed, but that was a matter of fairly straightforward professionalism.

2. *The failure of the Fregean consensus*

Famously, the work of Saul Kripke (1972, 1979, 1980) and others overturned the consensus. Proper names in natural language have no Fregean senses, at least of anything like the kind traditionally assumed by Fregeans. A more promising approach to the semantics of both names and indexicals treated them as *directly referential*: such an expression contributes *only* its ordinary referent and not also a sense to the proposition expressed by a sentence in which it occurs (Kaplan 1989). But the direct reference view gives new menace to Frege puzzles, since it seems to make the relevant substitutions truth-preserving for names and other directly referential terms, though not for definite descriptions. A long and complicated debate ensued, and still rumbles on, about the semantics and pragmatics of attitude ascriptions. I will not attempt to summarize all the moves and counter-moves, though I will sketch some reasons why Fregeanism failed to fulfil its initial promise. My main aim in this paper is to consider some implications of anti-Fregeanism for epistemology, with respect to both general methodology and specific epistemological theses.

Frege puzzles gave Fregeanism a very substantial head start over direct referentialism, of which Fregeans proved unable to take much advantage. What gradually became clear was that, even where it seems most promising to associate an expression with a mode of presentation of its referent, that mode does not play the semantic role in natural language that Fregeanism would lead one to expect.

An example is the first person pronoun. For Fregeans, the indexical 'I' is presumably associated with the mode of presentation of oneself to oneself *as oneself*, which we can call the *first-personal mode of presentation*. Now consider my ascription of a belief to you:

- (1) You believe that I was born on 6 August 1955.

Background: I was indeed born on that date; for the sake of the example, I will assume that you were not. Let p be the proposition which I say that you believe when I utter (1). Since ‘I’ occurs in (1) firmly within the ‘that’-clause, on the most straightforward Fregean approach it contributes a mode of presentation to p . Since ‘I’ is associated with the first-personal mode of presentation, it presumably contributes that mode of presentation to p . Thus, in uttering (1), I say that you believe p , where p is the proposition made up of the first-personal mode of presentation and a mode of presentation of something like the property of having been born on 6 August 1955. But, in your beliefs, the first-personal mode of presentation picks out *you*, not me, as the referent. Thus, in uttering (1), I end up attributing to you a false belief about your birthday, not a true belief about mine. That is absurd. Such a reading is just not available for the indirect speech ascription (1), by contrast with a direct speech ascription such as (2):

(2) You accept ‘I was born on 6 August 1955’.

Naturally, Fregeans can postulate various more convoluted readings of my utterance of (1), on which I am saying in effect that you have a belief in some non-first-personal proposition q suitably related to the first-personal proposition which *I* express by the sentence ‘I was born on 6 August 1955’, while not myself expressing q . But the supposed availability of such readings is not to the point, which is rather the *unavailability* of the absurd reading generated by the flat-footed application of the Fregean approach, on which I am attributing to you a false belief about your own birthday. As a normal speaker of English, I cannot hear such a reading, no matter how hard I try. I do not hear it only to exclude it immediately on pragmatic grounds; I just do not hear it in the first place. Something is wrong with an approach that gets anywhere near such a reading.

Although philosophers from Descartes on are used to giving the first person very special treatment, there is no evidence that the semantics of natural language affords it any such privilege. Linguistically, it would be quite implausible to dismiss examples like (1) as a special case.³

Of course, such examples will not silence Fregeans; nothing will. But they do suggest that the appeal to modes of presentation in semantics is much less natural, much less in tune with the workings of natural language and the needs of communication, than it first seemed.

We may well have to live with the conclusion that, despite appearances, the substitutions in Frege puzzles are truth-preserving. The main focus of this paper is on the epistemological consequences of that moral. I will use some ideas in Kripke’s classic paper ‘A Puzzle about Belief’ (1979, 1988) as a starting-point from which to explore and extend a non-Fregean approach to Frege puzzles, though I will intersperse some more comments about Fregean approaches. In the closing sections, I will discuss some implications for issues about evidence and subjective or epistemic probability.

‘A Puzzle about Belief’ is most famous for its examples, especially Pierre, who asserts ‘Londres est jolie’ but denies ‘London is pretty’, and Peter, who does not realize that the pianist Paderewski is the statesman Paderewski. Here we will be more concerned with some of the more general themes in Kripke’s discussion.

3. Frege puzzles and synonymy

Kripke emphasizes that Frege puzzles can arise for expressions of many kinds, even when the two expressions are normally regarded as *synonymous*, not just as co-referential. Thus one cannot simply diagnose Frege puzzles as arising when two terms have the same reference but different meanings, for by normal standards they can arise even for two terms with the *same* meaning. Of course, whether meaning is anything more than reference is itself contested in the semantic debate, but that is not our present concern. Even a direct referentialist might hold that co-reference varies with context while synonymy requires co-reference across all contexts; thus ‘Timothy Williamson’ and ‘I’ are co-referential in my context, but not synonymous, because they are not co-referential in your context.

One of Kripke’s examples concerns the synonyms ‘furze’ and ‘gorse’, which are simply two terms for the very same kind of shrub (1988: 134). The two words may have originated in different dialects of English. The shrub itself changes in appearance from one season to another, sometimes having dull brown needles and no flowers, sometimes bright green needles and yellow flowers. There is no *general* correlation between how the shrub appears and which word is applied, though for accidental reasons there may be such a correlation in how a particular speaker applies the words.

Suppose that Penny learns the term ‘furze’ somewhere by being shown various samples with bright green needles and yellow flowers, and learns the term ‘gorse’ separately somewhere else by being shown various samples with dull brown needles and no flowers. By normal linguistic standards, she understands both terms. Nevertheless, she may be in no position to know that the two terms co-refer. In that case, if one uses both terms with their normal meanings and happens to know that they co-refer, one may still feel tempted to describe Penny’s situation by asserting (3) and denying (4):

(3) Penny believes that furze is furze.

(4) Penny believes that furze is gorse.

But the guiding principle of formal semantics is *semantic compositionality*, according to which the meaning of a complex expression is determined by the meanings of its simpler constituents and the way in which they are put together. Since (3) and (4) are put together in the same way out of corresponding words with the same meanings, in the absence of semantic ‘funny business’ compositionality requires (3) and (4) to have the same meaning too. But since meaning and context determine truth-value, it follows that they also have the same truth-value in any given context, such as a context in which one both asserts (3) and denies (4). Thus such a combination can hardly be the right way to describe Penny’s situation.

One might worry that the repetition in (3) and lack of repetition in (4) makes a structural difference between (3) and (4). That would be insufficient to block the argument from a standard version of semantic compositionality, but we can anyway finesse the issue by using a pair without repetition, such as (5) and (6):

(5) Penny believes that furze has yellow flowers.

(6) Penny believes that gorse has yellow flowers.

There is the same temptation to assert (5) and deny (6), but there is also the same semantic reason as before to resist that temptation.

What kind of ‘funny business’ could block the arguments from semantic compositionality? In principle, verbs like ‘believe’ could introduce some sort of covert sensitivity to context, so that ‘S believes that p’ is true only if S assents to the proposition expressed by ‘p’ as presented in some contextually relevant way. Such an account permits contexts in which (4) differs in truth-value from (3), and (6) from (5), without violating semantic compositionality. For example, when the speaker is using the choice between ‘furze’ and ‘gorse’ to mark a contrast, the truth of (5) might require Penny to assent to the proposition that furze has yellow flowers under the guise of the sentence ‘Furze has yellow flowers’, while the truth of (6) requires her to assent to the same proposition under the guise of the sentence ‘Gorse has yellow flowers’.⁴

However, such contextualist hypotheses look *ad hoc*. Normally, just one of ‘furze’ and ‘gorse’ is used in a given conversation, and, although the example made Penny a native English speaker, the truth of (5) or (6) does not *require* her to know a word of English; she could use a natural kind term in her own language for the shrub. More radically, nothing in the semantics of English requires Penny to have a language of any kind for (3)-(6) to be true. Imagine a species of languageless animals whose diet consists solely of gorse (like pandas with bamboo shoots); they need and have a recognitional capacity for gorse, to which the belief that gorse has yellow flowers is crucial. Various forms of functionalism about the metaphysics of belief allow such a case; to exclude those theories in the philosophy of mind seems to be no business of the semantics of English. But then, if the truth-conditions of belief ascriptions are normally indifferent to the guise (if any) under which a believer believes a proposition, it seems unlikely that natural languages would have a special semantic mechanism waiting to spring into action just to resolve Frege puzzles and a few related difficulties. Theoretically, a more explanatory account of Frege cases would be derived from more general principles needed far beyond Frege cases, rather than relying on semantic structure postulated *ad hoc* just to handle those very cases.

Grice’s category of *conversational implicature* has the requisite generality, since he combines quite general principles of conversation with simple semantic assumptions to predict context-sensitive conversational implicatures (Grice 1989). Indeed, in some contexts, a speaker might well refrain from uttering (4) because uttering (4) would have the false conversational implicature that Penny would assent to ‘Furze is gorse’. In the same context, the speaker might also utter (5) rather than (6) because uttering (6) would have the false conversational implicature that Penny would assent to ‘Gorse has yellow flowers’. That is compatible with (4) and (6) semantically expressing true non-metalinguistic propositions in that context, which do not entail those metalinguistic conversational implicatures. In Nathan Salmon’s terminology (1986), the metalinguistic information is pragmatically imparted, not semantically encoded.

Unfortunately, even granted that there are such conversational implicatures, they do not explain all the phenomena in Frege cases. For speakers do not merely refrain from asserting (4) and (6); they may actively *deny* (4) and (6), and regard them as false. Normally, if uttering a sentence would have a false conversational implicature, that does not justify one in uttering the negation of that sentence. If saying ‘The Professor is sober this morning’ would have the false conversational implicature that the Professor is often drunk, that does not justify one in saying ‘The Professor is not sober this morning’. On an anti-contextualist anti-Fregean semantics, (4) and (6) are true in the example, and whoever denies them speaks falsely. Thus some sort of error theory is needed.⁵ Even if the error involves some confusion between conversational implicatures and truth-conditional consequences, that confusion would need to be explained, since it is untypical of conversational implicatures.

If anti-contextualist anti-Fregeans can resort to error theories, so can contextualist anti-Fregeans.⁶ However, positing *both* contextualist *and* error-theoretic mechanisms to explain the data seems unattractively uneconomical. If one has to posit an error-theoretic mechanism anyway, why not let it do all the work, and avoid the need to posit a contextualist mechanism as well? Contextualist anti-Fregeans might respond that a principle of charity in interpretation demands that we posit as little error as possible, and so explain as much of the data as we can with the contextualist mechanism. But if the errors are systematic, not just random performance errors, a well-developed error theory will posit a *specific* mechanism to explain them, which should indicate how widespread we can expect the errors to be. Thus we will need to know more about the putative error mechanisms before adjudicating the issue. Section 6 will discuss such mechanisms in detail.

Faced with these difficulties, some philosophers deny that cases like Penny’s can arise. One form of denial is to insist that Penny is not fully competent with the terms ‘furze’ and ‘gorse’, perhaps on the grounds that her understanding of them involves deference to expert botanists. That is not a promising strategy, for several reasons. First, the original description of the case did not mention semantic deference or a recognized scientific community, and does not require such elements. It simply involves some people being better than others at recognizing particular natural kinds, which happens in any community and does not entail semantic deference. Second, the strategy does not vindicate the observed combinations at issue, asserting (3) and (5) while denying (4) and (6). For if Penny’s alleged lack of full competence with the terms is relevant at all, it undermines reading off what she believes from her use of those terms; but we are happy to assert (3) and (5) *on the basis of* such reading off. Third, more generally, if having attitudes about a natural kind requires a perfect recognitional capacity for that kind, virtually no one has attitudes about any natural kind. Given that imperfect recognitional capacities suffice, Frege puzzles like this cannot be excluded.

Another form of denial is to insist that Penny has her own personal senses for ‘furze’ and ‘gorse’, which are different from each other and from ours. That too is an unpromising strategy. It rests on a misunderstanding of how attitude ascriptions work in natural language. When we make such ascriptions, the words in the ‘that’-clause mean what *we* mean by them, not what the subject to whom we are ascribing the attitude does. That already emerged in the discussion of the role of ‘I’ in sentence (1). Similarly, if Penny happened not to know the word ‘yellow’, which is quite consistent with the original scenario, that would not make

‘yellow’ in (5) and (6) *meaningless*; at worst, (5) and (6) might turn out to be false. Nor do we somehow try to make ‘furze’ and ‘gorse’ in (3)-(6) mean what Penny personally means by them. After all, we do not know exactly what she means by them. If we simply defer to her for their meanings in these sentences, we undermine the individualistic conception of meaning on which the objection initially relied, with its talk of Penny’s personal senses. Anyway, vicariously using the agent’s meaning is not an option once we generalize over the agent parameter: for example, when one asks ‘How many people know that gorse has yellow flowers?’, the embedded sentence ‘Gorse has yellow flowers’ does not have someone else’s meaning.

What matters for what (3)-(6) mean in our mouths is what ‘furze’ and ‘gorse’ mean in our mouths; they are synonymous in our mouths, and remain so when we utter (3)-(6). Thus denying that Penny uses the words ‘furze’ and ‘gorse’ in their normal English senses is not just implausible; like the previous strategy for denial, it fails to vindicate the combinations at issue, asserting (3) and (5) while denying (4) and (6).

Of course, we can consider metalinguistic variants of the sentences at issue. For example, instead of (5) and (6) we consider:

(5m) Penny believes that the sentence ‘Furze has yellow flowers’ expresses a truth.

(6m) Penny believes that the sentence ‘Gorse has yellow flowers’ expresses a truth.

There is no Frege puzzle here, because the metalinguistic beliefs are about distinct sentences. Even if (5m) is true, it obviously does not follow that (6m) is true. But that does not explain the Frege puzzle with (5) and (6). After all, (5m) and (6m) are not in general good paraphrases of (5) and (6) respectively. In a quite different scenario, where ‘Penny’ refers to someone who knows no English, (5m) and (6m) are typically false, while (5) and (6) may easily be true—for example, she may be an expert botanist, who would express the knowledge ascribed in each of (5) and (6) using the same word for the shrub in her own language.

It is hard to avoid the conclusion that, read strictly and literally, (3) has the same truth-value as (4), and (5) the same truth-value as (6). Analogous considerations apply when ‘believes’ is replaced by ‘knows’ in (3)-(6). More generally, epistemologists should not assume that Frege puzzles can be handled in some vaguely Fregean or contextualist way. We must be ready to be much more revisionary in our treatment of the initial judgments.

4. *Frege puzzles from the inside*

A Fregean might concede that the sense-reference distinction is ill-adapted to the semantics of natural language, which is tailored to the needs of inter-personal communication, but still insist that it is just right for the needs of individual thought, including reflection on the epistemic status of one’s own beliefs. On this view, the problems we have in saying what Pierre, Peter, or Penny believes are fundamentally problems in using a public natural language to describe a private cognitive perspective. This concession would have been

unwelcome to Frege himself, for he emphasized that thoughts (the senses of declarative sentences) can be part of the common heritage of humankind, as mathematical theorems are. But we can still consider individualistic Fregeanism in its own right.

By itself, the restriction to first-person ascriptions is insufficient. When Pierre later learns that ‘Londres’ and ‘London’ refer to the same city, and when Penny later learns that ‘furze’ and ‘gorse’ refer to the same shrub, they can wonder what beliefs they had before the discovery, and find the case just as puzzling to describe as we outsiders do. To avoid such problems, what is needed is at least a restriction to first-person *present-tense* ascriptions. But that is still not enough. For Pierre and Penny can entertain the possibility of co-reference even before they know or believe that it obtains. Suppose that Penny is agnostic as to whether ‘furze’ and ‘gorse’ co-refer. She can think to herself: ‘Perhaps furze *is* gorse; in that case, since I believe that furze has yellow flowers, do I also believe that gorse has yellow flowers?’ Pierre can ask himself analogous questions. Even in the first person and present tense, such questions remain puzzling.

The Fregean needs to go further, by postulating new attitudes to Fregean thoughts, rather than using the natural language apparatus for attitude ascription. Let the underlined expression ‘e’ refer to the supposed Fregean sense which the sentence ‘e’ has for Penny right now, in her agnostic state. The idea is that furze and gorse are distinct, because ‘furze’ and ‘gorse’ have different senses for Penny in her agnostic state. Thus furze has yellow flowers and gorse has yellow flowers are distinct Fregean thoughts, for the sense of a complex expression is supposed to be a structured entity built of the senses of its simpler constituents; this is a strong form of semantic compositionality for senses. Let belief_s be the attitude analogous to belief which Penny has to some Fregean thoughts.⁷ The idea is that Penny believes_s furze has yellow flowers but does not believe_s gorse has yellow flowers; there is no paradox in her believing one Fregean thought and not another. She can use this apparatus to reflect on her own beliefs and their epistemic status. Presumably, she knows both that she believes_s furze has yellow flowers and that she does not believe_s gorse has yellow flowers, from which she can infer that furze and gorse are distinct, by compositionality. At this level, Frege puzzles are easily resolved.

When Penny later learns that ‘furze’ and ‘gorse’ co-refer, they will presumably come to have the same sense for her. Since they have distinct senses right now, at least one of the words will change its sense for her; by the symmetry of the situation, both will. Consequently, she cannot now rely on underlining (or her mental equivalent of it) to describe her future thoughts, since underlining is tied to the senses expressions have for her right now. This already indicates a serious limitation of the envisaged Fregean apparatus for purposes of her epistemic deliberation. In considering what to believe, she must think about what her belief state will be after a potential change. For example, when she considers whether to accept ‘Furze is gorse’ and ‘Gorse has yellow flowers’, she must think about what senses they will have for her after she has accepted them, which depends on the new sense ‘furze’ and ‘gorse’ will have for her; but it is not the sense of any of her current words, and it is unclear how she can even entertain it until she has made the very belief change whose merit she is currently trying to assess. By contrast, if Penny works in natural language, these difficulties do not arise: if she makes the envisaged change, she will believe that furze is

gorse, but presumably will not believe, furze is gorse, since she will no longer have words with the senses furze and gorse.

A further concern for the Fregean apparatus can be explained with a variant of Penny's case. Pat acquired 'furze' and 'gorse' separately as natural kind terms, in a completely normal ostensive way, though in circumstances she no longer remembers. She applies 'furze' by using a reliable recognitional capacity, and she applies 'gorse' by using an exactly similar and so equally reliable recognitional capacity. For her, the *only* differences between 'furze' and 'gorse' in associated descriptions and recognitional capacities are metalinguistic: 'furze' and 'gorse' are different words. However, she does not treat them completely interchangeably, for she has a slight worry that they may not co-refer: she worries that she may have been introduced to 'furze' as a word for one kind of shrub and to 'gorse' as a word for a different kind of shrub, though they are so similar in appearance that she cannot recognize the difference. In fact, that is not what happened; she was shown the same kind both times. Still, her doubt need not be neurotic: she may have had ample experience of lookalike species. Thus she is slightly more confident of 'Furze is furze' than of 'Furze is gorse'. In Fregean terms, she is slightly more confident, of furze is furze than of furze is gorse (here the underlining is for Pat's senses, not Penny's). Consequently, furze and gorse are different senses. But 'furze' and 'gorse' do not differ in meaning for Pat, unless the difference between the words themselves counts as a difference in meaning.

Although Pat was described as thinking about the words themselves, that is not crucial to the example. Her slightly greater confidence in response to 'Furze is furze' than in response to 'Furze is gorse' suffices for the argument, and does not require her to think metalinguistically.

What is wrong with building the words themselves into their Fregean senses for a given thinker? Such senses do not fit the usual conception of Fregean senses, but that is not the main problem. Instead, it is an issue of motivation. The required theoretical apparatus of individualistic Fregean senses is elaborate, unclear, and ill-developed. It is not useful for natural language semantics. Nor does it look useful for epistemology, given the problem of incommensurability of senses between different cognitive states. After all, a significant part of epistemology concerns the communication of knowledge across such differences—by memory, from earlier to later states of the same individual, and by testimony, from states of one individual to states of another. Learning itself—the acquisition of knowledge—constitutes a change in cognitive state. What gets preserved through such transactions is likely to be quite coarse-grained.

Once the senses of words are individuated in terms of the words themselves, the senses risk being redundant. We already have the words themselves, and their meanings in natural language. The senses add another level of theoretical entities with no clear explanatory value. We can make the distinctions we need without them.

5. *The necessary a posteriori and the contingent a priori*

As Kripke recognizes, his revisionary treatment of Frege puzzles affects even some of his own signature doctrines in *Naming and Necessity*, especially concerning the necessary *a*

posteriori and the contingent *a priori* (1988: 135, 147n44). For example, on Kripke's view of the necessary *a posteriori* as often expounded, we know the necessary truth that Hesperus is Phosphorus only *a posteriori*, so while (7) is true, (8) is false:⁸

(7) We know *a priori* that Hesperus is Hesperus.

(8) We know *a priori* that Hesperus is Phosphorus.

This is just another Frege puzzle. Kripke is open to the possibility that, on the correct semantics, (7) and (8) have the same truth-value (with 'Hesperus' and 'Phosphorus' treated as directly referential proper names).

Kripke suggests a metalinguistic fall-back for the problematic claim 'It was once unknown that Hesperus is Phosphorus': 'we can still say that there was a time when men were in no epistemic position to assent to "Hesperus is Phosphorus" for want of empirical information, but it nevertheless expressed a necessary truth' (1988: 135). He explains: 'I was aware of this question by the time "Naming and Necessity" [Kripke 1972] was written, but I did not wish to muddy the waters further than necessary at that time'.

For the contingent *a priori*, imagine a variant of Kripke's scenario in which we fix the reference of the rigid, directly referential term 'metre' by the non-rigid description 'the length of stick S', and we fix the reference of the rigid, directly referential term 'metre*' by the non-rigid description 'the length of stick S*'. As it happens, and unknown to us, the two sticks S and S* are exactly the same length. Consider these four statements:

(9) We know *a priori* that stick S is one metre long.

(10) We know *a priori* that stick S* is one metre* long.

(11) We know *a priori* that stick S is one metre* long.

(12) We know *a priori* that stick S* is one metre long.

On Kripke's view of the contingent *a priori* as often expounded, in all four cases, the embedded proposition about the length of the stick is contingent; (9) and (10) are true, but (11) and (12) are false. For (9) and (10) follow the definitional connections, while (11) and (12) cut across them. But 'metre' and 'metre*' are directly referential terms for the very same length; on the revisionary view of Frege puzzles, substituting one for the other in such contexts preserves truth-value. Thus (11) has the same truth-value as (9) and (12) the same truth-value as (10). Moreover, since (9) and (10) uncontentiously have the same truth-value as each other because the two cases are exactly parallel, (9)-(12) all have the same truth-value. That is quite contrary to the standard account.⁹

As with the necessary *a posteriori*, metalinguistic fall-backs may still be available, for example about people in no epistemic position to assert 'Stick S is one metre* long' or 'Stick S* is one metre long' for want of empirical information. More generally, Kripke comments that when he wrote his (1972), 'I regarded the distinction between epistemic and

metaphysical necessity as valid in any case and adequate for the distinctions I wished to make' (1988: 147n44).¹⁰

A problem for Kripke's fall-backs is that both epistemic necessity and empirical informativeness are also subject to problems of substitutivity:

- (13) It is epistemically necessary that Hesperus is Hesperus.
- (14) It is epistemically necessary that Hesperus is Phosphorus.
- (15) The information that Hesperus is Hesperus is empirical.
- (16) The information that Hesperus is Phosphorus is empirical.

We might expect (13) and (16) to be true and (14) and (15) to be false. But if substituting 'Phosphorus' for 'Hesperus' preserves truth from (7) to (8), as Kripke holds it may do, it should also preserve truth from (13) to (14) and from (16) to (15). Thus 'epistemic necessity' and 'empirical information' will have to be reconstructed in some other form if they are to play the required role in proofing the categories of the necessary *a posteriori* and the contingent *a priori* against substitutivity problems. A more thoroughly metalinguistic approach may be needed.

To some, Kripke's appeal to the distinction between epistemic and metaphysical necessity will suggest two-dimensional semantics, for example as developed by David Chalmers (2006), with one dimension epistemic and the other metaphysical. Chalmers conceives his semantics as Fregean in spirit, by contrast with two-dimensional semantics in the tradition of David Kaplan (1989), which is motivated by more purely linguistic considerations, treats proper names as directly referential, and has no epistemic dimension. For instance, Kaplan assigns exactly the same semantic properties to the names 'Hesperus' and 'Phosphorus', whereas Chalmers distinguishes them on the epistemic dimension of his semantics. In 'A Puzzle about Belief', Kripke holds open the possibility of a semantic theory such as Kaplan's with directly referential terms, and so cannot simply appeal to Chalmers' quite different approach.

In effect, for Chalmers a word has a Fregean sense (or 'narrow content') only relative to a given individual at a given time. The individualistic Fregean senses in section 4 might be fitted into the framework of Chalmers' semantics, but that gives a hint of the difficulties it faces. It also contrasts methodologically with Kripke's strong emphasis on the sharing of linguistic meaning, for instance as names are passed down reference-preserving historical chains.

A further concern for the use of Chalmers' framework to stabilize the distinction between epistemic and metaphysical modality is his appeal to *a priori* knowability in characterizing epistemic modality, given that the distinction between *a priori* and *a posteriori* knowledge is itself destabilized by substitution arguments, as in (7)-(12). Clearly, this is not the place for a detailed assessment of Chalmers' grand programme, with its speculative reductionist ambitions, though the problems raised in section 4 are very relevant.¹¹ If his Fregean programme were on the right lines, that would have many ramifications for

epistemology. In what follows, I will assume that his programme is *not* on the right lines, so we face the task of working out the epistemological consequences of an anti-Fregean approach.

Clearly, the anti-Fregean treatment of Frege puzzles has non-trivial consequences for the epistemological distinction between *a priori* and *a posteriori* knowledge. In particular, it raises concerns about the reliability of our pre-theoretic assessments of attributions of those types of knowledge. That leaves us with a picture significantly less clear than the one to be found in *Naming and Necessity*.¹²

6. Principles for belief ascription

In ‘A Puzzle about Belief’, Kripke suggests an explanation for our difficulties in knowing what beliefs to ascribe in Frege puzzles. He formulates three principles which he takes to guide our ascription of beliefs, but which jointly generate problematic consequences when applied to cases like that of puzzling Pierre. The first he calls the ‘*disquotational principle*’ (1988: 112-13):

(DP) *If a normal English speaker, on reflection, sincerely assents to ‘p’, then he believes that p.*

The second is a strengthened biconditional form of the disquotational principle (1988: 113):

(SDP) *A normal English speaker who is not reticent will be disposed to sincere reflective assent to ‘p’ if and only if he believes that p.*

The third he calls the ‘*principle of translation*’ (1988: 114); unlike (DP) and (SDP), it is not specifically about belief, or even propositional attitudes in general:

(TP) *If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language).*

Kripke stipulates that the schematic letter ‘*p*’ in (DP) and (SDP) ‘is to be replaced, inside and outside all quotation marks, by any appropriate standard English sentence’, which ‘is to lack indexical or pronominal devices or ambiguities that would ruin the intuitive sense of the principle’ (1988: 112-13). Kripke seems to understand both (DP) and (SDP) as implicitly generalized to other natural languages too, in particular to French. While admitting that (DP) may need further qualifications, he says of (DP): ‘Taken in its obvious intent, after all, the principle appears to be a self-evident truth’ (1988: 113).

When we apply (DP) and (TP) to Pierre’s linguistic behaviour in French and English, we are led to ascribe contradictory beliefs to him, because he sincerely and on reflection assents to a French sentence (‘Londres est jolie’) and an English sentence (‘London is not pretty’), where the latter is the negation of the English translation of the former: he both

believes that London is pretty and believes that London is not pretty. That is paradoxical because we may assume that, by normal standards, Pierre is eminently rational.

When we apply (SDP) and (PT) to Pierre's linguistic behaviour in French and English, we are led to make contradictory statements about Pierre's beliefs—unlike the previous case, where we just attribute contradictory beliefs to Pierre. For he is not reticent, and he is disposed to sincere, reflective assent to a French sentence ('Londres est jolie') but not to its English translation ('London is pretty').

Kripke argues that, since our normal practice of belief attribution suffices to generate the problem, given a mild principle of translation, with no appeal to substitutivity, it would be wrong-headed to blame the problem on the latter. Indeed, cases such as that of 'furze' and 'gorse' suggest that even the principle of translation is stronger than needed to generate the problem; a principle of synonymy within a single language would do instead. Kripke concludes his paper by saying that the puzzle cases lie 'in an area where our normal apparatus for the ascription of belief is placed under the greatest strain and may even break down' (1988: 136).

What would it be for 'our normal apparatus for the ascription of belief' to 'break down'? Kripke writes of it as an extreme outcome which 'may even' occur. Thus it must be something worse than just our finding ourselves unsure what to say about the puzzle cases, since that obviously does occur.

Is Kripke hinting that the very distinction between believing and not believing may not apply to such cases? That is what some people would expect if, for example, (SDP) had some sort of analytic status, making it quasi-definitional of 'believe'. But (SDP) is nothing like analytic. For instance, a congenital liar may believe that *p* without being at all disposed to sincere reflective assent to '*p*'; he may non-reticently deny '*p*' (he is still a normal English speaker in his linguistic capacities). Even (DP) may have counterexamples, unless 'sincerely' is tied too closely to the expression of belief for (DP) to be a helpful guide to belief ascription; if '*p*' is a complicated sentence with several negations, a normal English speaker might make a performance error in processing it and, even on reflection, sincerely assent to '*p*' although it in fact expresses the opposite of what he believes. Kripke himself expresses doubts on the matter: he writes of (DP) 'I fear that even with all this [qualification] it is possible that some astute reader—such, after all, is the way of philosophy—may discover a qualification I have overlooked, without which the asserted principle is subject to counter-example', and of (SDP) 'Maybe again the formulation needs further tightening, but the intent is clear' (1988: 113-14).

A more promising view of (DP) and (SDP) is that they are standard *heuristics* for belief ascription, rules of thumb which are comparatively quick and easy to use, but not perfectly reliable. An analogy is our visual system's use of colour boundaries as a fallible guide to the shapes of three-dimensional objects before us; camouflage deceives us by exploiting that heuristic. Most heuristics have such limitations.¹³ Kripke's puzzle cases bring out some limitations of (DP) and (SDP).

A much simpler basic heuristic may indeed underlie both (DP) and (SDP):

(BDP) *English speakers assent to 'p' if and only if they believe that p.*

For the qualifications about normality, reflection, sincerity, reticence, and dispositionality all look like exception-barring clauses inserted to guard against defeaters. (BDP) is easier to use than (DP) or (SDP): not only is it simpler, it provides an overt criterion—assent to a sentence—for belief in a proposition, while all the qualifications concern less easily observed matters (normality, reflection, sincerity, reticence, dispositionality) from the perspective of another person. Of course, our capacity to recognize the need for these qualifications shows that (BDP) does not *exhaust* our understanding of belief, but such defeasibility is typical of heuristics. As Kripke's uncertainty over the formulation of (DP) and (SDP) suggests, being a native speaker does not put one in a position to survey all possible ways for the heuristic to be defeated; as one tries to think of as many as one can, pre-theoretical reflection gradually becomes more theoretical. When we use a heuristic, either a qualified one like (DP) and (SDP) or an unqualified one like (BDP), we may easily be unaware of its merely heuristic status, or indeed of what principle we are using, if any.

Heuristics are very different from conversational implicatures. Unlike the latter, heuristics play a direct role in assessing statements as true or false. Also, conversational implicatures are more or less predictable on general social grounds, whereas heuristics are cognitive devices applicable to specific kinds of statement. Of course, heuristics may help to generate conversational implicatures. For instance, if John asks 'Does Penny believe that gorse has yellow flowers?' and Mary replies 'She believes that furze has yellow flowers', John may take Mary to imply that the literally correct answer to his question is negative, or at least that she does not know it to be positive, since otherwise the switch from 'furze' to 'gorse' would be pointless. But such conversational phenomena are secondary.

One limitation of (DP), (SDP), and (BDP) is that they are applicable only to the ascription of beliefs to speakers of a language. But we often ascribe beliefs to non-speakers, such as non-human animals and very young children, by observing their non-verbal behaviour. For example, to explain why the cat jumped into the bathtub, we assume that she believed (falsely) that it was empty. Frege cases arise for such creatures too. To explain why a cock robin in his own territory who sees himself in a mirror flies aggressively at the mirror, we assume that he believes (falsely) that he is a rival. To keep track of the robin's mental states, we may distinguish guises: he believes that he [under a visual demonstrative guise] is a rival, but does not believe that he [under a self-relating guise] is a rival. That is doubtless not *exactly* how the robin thinks, but it may be a decent first approximation. Similar examples could be given for very young children. Like non-human animals, they can be tricked by Frege cases into false belief without being able to grasp what false belief is.¹⁴

Of course, we can ascribe beliefs to older children and adult humans, on the basis of either linguistic or non-linguistic behaviour. Which sort of heuristic is more reliable, and which is more convenient, depends on the case. If we lack the time or opportunity to observe someone's non-linguistic behaviour, but want to know whether they believe that *p*, just asking them '*p*?' is often a good way to find out.

A more general limitation of Kripke's account in the paper is that it is specific to belief, although Frege puzzles arise just as much for other attitudes, such as knowledge. It is not obvious how to generalize (DP), (SDP), or (BDP) to those other attitudes. Still, (DP), (SDP), and (BDP) are a start. Given that knowledge entails belief, whenever (SDP) indicates the absence of belief, it also indicates the absence of knowledge. On the positive side,

someone's assent to 'p' might be taken as a highly defeasible sign that they know that p; our knowledge that it was not the case that p would be a salient defeater. At a more general level, the same mindreading capacity used for knowledge ascription is also used for belief ascription (Nagel 2013).

For any propositional attitude ϕ , if one wants to know whether a normal speaker ϕ s that p, in many circumstances a quick and moderately reliable way to find out is by asking them 'Do you ϕ that p?' We can call this the 'Just Ask' heuristic. It is not quite a generalization of (DP), (SDP), or (BDP) for in the case of belief it tells one to ask 'Do you believe that p?', whereas for (DP), (SDP), and (BDP) one would ask simply 'p?' The questions are not equivalent. If you ask honest agnostics 'Do you believe that there is a god?' they will answer 'No' (agnostics are not theists); by contrast, if you ask them 'Is there a god?' they will not answer 'No' (agnostics are not atheists), but instead 'I don't know', or the like. Nevertheless, the 'Just Ask' heuristic has a key feature in common with (DP), (SDP), and (BDP): all these tests are sensitive to differences between the *sentences* 'p' and 'q', even if the *propositions* that p and that q are identical. Consequently, they may help explain our troubles in Frege cases for any attitude. If the propositions that p and that q are identical, a rational person may still give conflicting answers to the questions 'Do you ϕ that p?' and 'Do you ϕ that q?' Heuristics for attitude ascription with this feature are *language-sensitive*.

By contrast, the factivity principle 'If not-p, the agent does not know that p' is language-insensitive. For if the propositions that p and that q are identical, then not-p if and only if not-q, so the principle never yields conflicting answers to the question 'Does the agent know that p?', though we may of course misapply the principle if we are confused about the relation between 'p' and 'q'.

When ϕ ing is a factive attitude, the question 'Do you ϕ that p?' normally presupposes that p. For example, the question 'Do you regret that you never told him what you thought of him?' presupposes that the addressee never told the relevant man what they thought of him. The fact itself is taken for granted; the question concerns the addressee's affective attitude to the fact. That in itself poses no problem for the 'Just Ask' heuristic.

However, when ϕ ing is a more purely cognitive factive attitude, such as knowing, seeing, or remembering, often one asks whether someone ϕ s because one wants to *learn* the fact at issue; one is in no position to ask 'Do you ϕ that p?' because one does not yet know whether p. In such cases, a slightly more complex variant of the 'Just Ask' heuristic is needed. Unless one already knows that the train stops at Ardlui, one does not ask 'Do you know that the train stops at Ardlui?' (which might elicit the answer 'I do now'); one asks 'Do you know whether the train stops at Ardlui?' If one gets the rather uncooperative answer 'I do', leaving one still not knowing whether it stops there, one can follow up with 'So *does* it stop there?' More generally, the variant of the 'Just Ask' heuristic has one ask 'Do you ϕ whether p?', with 'p?' as the potential follow-up to an uncooperatively minimal positive answer. Similarly, one might ask 'Do you remember whether I locked the door?' or 'Can you see whether the light is on?' Rather than directly asking the question one is really interested in, one checks whether one's interlocutor is in a position to answer it. This variant of 'Just Ask' is, of course, equally language-sensitive.

'Just Ask' and its variants are more general than (DP), (SDP), and (BDP) because one can apply the former but not the latter to any attitude one can articulate. Obviously, they are

not fully reliable, because even a speaker who is trying to be honest may lack self-knowledge or be self-deceived. Nevertheless, they may be better than the available alternatives. One can also use ‘Just Ask’ in the past tense to probe the speaker’s past mental states, by asking questions of the form ‘Did you believe/hope/fear that ... ?’, ‘Did you already know that ...?’, ‘Did you remember that ... ?’, ‘Could you see whether ... ?’ By contrast, (DP) and (SDP) target only the speaker’s *present* beliefs; if ‘p’ is in the past tense, they target the speaker’s present beliefs about the past.

The primary use of these language-sensitive heuristics is ‘online’, to find out the attitudes of a living person with whom one can communicate. But of course that is not what readers of Kripke’s article are doing when they wonder what Pierre in the story believes. If they are applying (DP) or (SDP), they are doing it ‘offline’, in their imaginations. Such offline uses of heuristics for attitude ascription are not at all uncommon; we make them all the time when reading novels. In the case of Pierre, we need not imagine him talking to us or anyone else; we can simply imagine him saying ‘Londres est jolie’ or ‘London is not pretty’ to himself; (DP) and (SDP) are still applicable.

When we read works of analytic epistemology, we are often expected to do something similar, although the attitude at issue may be knowledge or justified belief rather than plain belief. The reader is asked to imagine fictional cases and, within the fiction, to make positive or negative attitude ascriptions. The content of the putative attitude is normally presented to the reader as expressed by a sentence. Naturally, we will use offline whatever heuristics we have for ascribing the attitude at issue, which will often involve a language-sensitive heuristic. This is a potential source of error in our use of the case method in epistemology. Of course, if the fictional scenario involves an obvious Frege case, experienced philosophers are likely to be on the alert for associated problems, but in some cases even the unmentioned possibility of a Frege case may make problems. For instance, when one considers the assertive use of Moore-paradoxical sentences of the form ‘I falsely believe that p’, it may not be immediately obvious that, given anti-Fregeanism, such uses can be legitimate in Frege cases (Crimmins 1992).

Our reliance on imperfectly reliable heuristics in assessing attitude ascriptions also threatens the standard methodology for studying the semantics of such ascriptions. The assessments of attitude ascriptions with respect to hypothetical cases are standardly treated as the central data for that study. Normally, a semantic theory of the truth-conditions of attitude ascriptions is expected to explain the data by vindicating the assessments, predicting their correctness. But if the assessments are the outputs of imperfectly reliable heuristics, then they may be *incorrect*, in which case a semantic theory should not predict their correctness. Thus the literature on the semantics of propositional attitudes may have fallen into the trap known in the natural sciences as *overfitting*, producing increasingly complicated theories to fit unreliable data (Forster and Sober 1994). If the errors result systematically from reliance on comparatively simple heuristics, semanticists need to know what those heuristics are, so that they can be taken into account in a correspondingly systematic way. We can still expect a semantic theory to treat the heuristics as charitably as possible, by not imputing errors needlessly, but to do that we must first understand how the heuristics work. Thus a more sophisticated methodology is called for, with a more critical attitude to the data.

The next section explores in detail some further consequences of an anti-Fregean treatment of Frege cases. For simplicity, I will assume that the anti-Fregean can explain all the systematically recalcitrant data in Frege cases as effects of reliance on imperfectly reliable heuristics for attitude ascription, and so has no need to invoke a contextualist semantics. Thus our anti-Fregean will henceforth be an anti-contextualist anti-Fregean.

7. Evidence

Normally, whether it is rational for one to believe a proposition p is sensitive to how p is related to one's total evidence. Thus, if one's evidence can be *non-transparent*, in the sense that what it seems to one to be is distinct from what it is, one may be in no position to know how p is related to one's total evidence, and so in no position to know whether it is rational for one to believe p . Similarly, whether it is rational for one to do an action A is normally sensitive to how A is related to one's total evidence. Thus, if one's evidence can be non-transparent, one may be in no position to know how A is related to one's total evidence, and so in no position to know whether it is rational for one to do A . Even the cleverest agents might be in such circumstances, whether they know it or not. Such possibilities matter for how we understand the normative claims of rationality.

But can evidence be non-transparent? Anti-Fregeanism about Frege puzzles suggests that it can. For the time being, we may assume that evidence consists of propositions, since it can be *incompatible* with hypotheses and stand in other such propositional relations (Williamson 2000: 194–200). In particular, we may assume that it can include such propositions as that furze has yellow flowers. Recall Penny from section 3. We may assume that her evidence includes that furze has yellow flowers; after all, she observed that furze has yellow flowers when she was introduced to the word 'furze'. If you ask her 'Does your evidence include that furze has yellow flowers?', on reflection she will sincerely answer 'Yes'. If you ask her 'Does your evidence include that gorse has yellow flowers?', on reflection she will sincerely answer either 'No' or 'I don't know', in part depending on whether *she* is a Fregean or an anti-Fregean. But, if anti-Fregeanism is in fact correct, for her evidence to include that furze has yellow flowers just is for her evidence to include that gorse has yellow flowers. Thus her evidence includes that gorse has yellow flowers. Consequently, her failure on reflection to answer 'Yes' to the question 'Does your evidence include that gorse has yellow flowers?' looks like a case of non-transparency. After all, she understands the question by normal linguistic standards.

We must go carefully here. Consider (17) and (18):

- (17) Penny knows that her evidence includes that furze has yellow flowers.
- (18) Penny knows that her evidence includes that gorse has yellow flowers.

In the scenario, (17) may well be true. One can know something by knowing it under one linguistic guise without knowing it under all its linguistic guises. But, given anti-Fregeanism,

(17) and (18) have the same truth-value. Thus (18) is true too. Since Penny *knows* that her evidence includes that gorse has yellow flowers, where is the alleged non-transparency?

Unfortunately, (18) does not resolve the problem of non-transparency. For Penny *also* has a psychologically real attitude of *doubt* towards the same proposition, where doubt is understood as an active attitude, not simply as absence of knowledge. Consider (19) and (20):

(19) Penny doubts that her evidence includes that furze has yellow flowers.

(20) Penny doubts that her evidence includes that gorse has yellow flowers.

In the scenario, (20) is true, because she doubts the proposition under the guise of sentence (22), even though she does not doubt it under the guise of sentence (21):

(21) My evidence includes that furze has yellow flowers.

(22) My evidence includes that gorse has yellow flowers.

Given anti-Fregeanism, (19) and (20) have the same truth-value, so (19) is also true. But Penny knows the very truth she doubts, because she knows it under the guise of (21), even though she does not know it under the guise of (22).¹⁵ In such a case, once we eliminate the relativization to guises, the non-transparency takes the form of the presence of doubt, rather than the absence of knowledge.

In some variants of the scenario, Penny believes that furze and gorse are distinct kinds. In that case, presumably, she will on reflection sincerely answer ‘No’ to the question ‘Does your evidence include that gorse has yellow flowers?’. Consider (23) and (24):

(23) Penny believes that her evidence does not include that furze has yellow flowers.

(24) Penny believes that her evidence does not include that gorse has yellow flowers.

In this scenario, (24) is true because Penny believes the proposition under the guise of the negation of sentence (22), even though she does not believe it under the guise of the negation of sentence (21). Given anti-Fregeanism, (23) and (24) have the same truth-value, so (23) is also true. Thus Penny believes contradictory propositions about her evidence. She believes that it includes that furze has yellow flowers (under the guise of sentence (21), because she knows it under that guise and knowledge entails belief), but she also believes that it does *not* include that furze has yellow flowers (under the guise of the negation of sentence (22)). The non-transparency takes the form of mutually contradictory beliefs as to what her evidence includes.

Even where there is no Frege case, the *epistemic threat* of a Frege case may result in non-transparency which may take the form of ignorance. For example, suppose that Dominic suspects that K2 (under that name) and Kangchenjunga (under that name) are the very same mountain. For all he knows, ‘K2’ and ‘Kangchenjunga’ co-refer. He knows that K2 is 8,611 metres high. He is not sure how high Kangchenjunga is, though of course he suspects that it

is 8,611 metres high. In fact, they are different mountains, and Kangchenjunga is slightly lower than K2. Thus, we may assume, his total evidence includes that K2 is 8,611 metres high but does *not* include that Kangchenjunga is 8,611 metres high. Consider (25):

- (25) Dominic knows that his evidence does not include that Kangchenjunga is 8,611 metres high.

In this scenario, given anti-Fregeanism, (25) is false. Everything Dominic knows is compatible with a scenario in which ‘K2’ and ‘Kangchenjunga’ co-refer and his evidence includes that Kangchenjunga is 8,611 metres high, because it includes that K2 is 8,611 metres high. Although he entertains the proposition that Kangchenjunga is 8,611 metres high, and his evidence does not include it, he is in no position to know that his evidence does not include it. Dominic is ignorant of the limits of his evidence.

Incidentally, not even the supposition that Fregeanism is true rules out this example of non-transparency, for it does not entail that Dominic *knows* that Fregeanism is true; thus everything he knows may be compatible with a scenario in which anti-Fregeanism is true, ‘K2’ and ‘Kangchenjunga’ are intersubstitutable in propositional attitude ascriptions, and his evidence includes that Kangchenjunga is 8,611 metres high, because it includes that K2 is 8,611 metres high. By itself, Fregeanism does not put Dominic in a position to know the limits of his evidence.

Despite all these problems, some anti-Fregeans may suspect that there is something right about the transparency of evidence. The preceding discussion makes one fall-back salient: to postulate transparency of evidence at the level of guises, instead of at the level of propositions. On this view, although it is not transparent to Penny whether her evidence includes the proposition that gorse has yellow flowers, and not even transparent to her whether it includes that proposition under the guise of the sentence ‘Furze has yellow flowers’ (since she doubts that ‘Furze has yellow flowers’ is a guise of the proposition that gorse has yellow flowers), it *is* transparent to her that her evidence includes whatever proposition is expressed by the sentence ‘Furze has yellow flowers’. More generally, the strategy would be to finesse the problem of non-transparency by doing as much of the epistemological work as possible with the guises themselves.

Kripke’s example of Paderewski the pianist and Paderewski the statesman raises immediate questions for the metalinguistic strategy. Non-transparency can arise even at the level of names and sentences. Similarly, as David Kaplan (1989) points out, it may be unclear whether two occurrences of ‘that’ as a perceptual demonstrative refer to the same object. Thus individuating guises by expression types in a natural language is insufficient.

Proponents of the strategy are likely to respond by going ever more psychological, perhaps treating the word types as mere proxies for underlying ‘mental files’.¹⁶ For example, Peter has one mental file associated with the phrase ‘Paderewski the pianist’ and another associated with the phrase ‘Paderewski the statesman’; each time we use ‘that’ non-anaphorically as a perceptual demonstrative, we open a new associated temporary mental file, and so on. Such mental files are reminiscent of the individualistic Fregean senses discussed in section 4, and raise related questions.

There may indeed be the postulated mental files, but they are not well-suited to rehabilitating the transparency of evidence, because they face their own Frege puzzles.¹⁷ After all, a paper file can have one label on the front and another on the back. I do not know whether I associate the English word ‘dog’ and the French word ‘chien’ with the same mental file or with two different but cross-referenced mental files, possibly with different contents. Nor is it clear to me how much my postulated mental file(s) contain. Similarly, Pat in section 4 may not know whether her words ‘furze’ and ‘gorse’ are associated with the same mental file or two different ones. The mere difference of words enables her to ask herself the non-trivial question ‘Is furze really gorse?’, irrespective of how the words are hooked up to mental files. Issues about the underlying organization of our mental filing system call for investigation by cognitive psycholinguistics; they are not settled by introspection. The problem is not with the specific metaphor of mental files, but with the more general strategy of postulating unconscious cognitive architecture in order to solve the problem of the non-transparency of evidence. Such architecture does not do the job it was called in to do.

A further problem for the attempt to do epistemology with propositional guises rather than propositional contents is that many epistemological relations are best understood at the level of content. For example, suppose that you are trying to determine whether a tree is dead by looking at it. The hypothesis to be assessed has a primarily linguistic guise: ‘This tree is dead’. The evidence by which it is to be assessed has a quite different, primarily visual format. Yet the latter may bear strongly on the former. At a first pass, we can put it in terms of a space of possible worlds: let *S* be the region where what you can see to be the case holds, and *D* be the region where the tree is dead; then most of *S* may be inside *D* (the hypothesis is very probable on the evidence), or most of *S* may be outside *D* (the hypothesis is very improbable on the evidence), or neither (the hypothesis is neither very probable nor very improbable on the evidence). That is a relation between the content (truth-conditions) of the evidence and the content (truth-conditions) of the conclusion. By contrast, if we compare the visual guise of the evidence with the linguistic guise of the hypothesis, they are quite disparate; the evidential relations cannot be understood at that level.

All this suggests that there is no level at which evidence or some proxy for it is transparent. Elsewhere, I have argued for the same conclusion on more purely epistemological grounds (Williamson 2000); here, my interest is in showing that it can also be reached by consideration of Frege puzzles.

Still, for modelling purposes, we can mitigate the problem in *ad hoc* ways (see Williamson 2017 for a general discussion of model-building in philosophy). For instance, to see how things look from the perspective of Penny or Pat, we can treat ‘furze’ and ‘gorse’ *as if* they were semantically independent, by allowing metaphysically impossible pseudo-worlds at which they are not coextensive, but which otherwise behave normally. Such worlds may later be epistemically ruled out for the agent by subsequently acquired evidence. That is not a semantic insight, for the words are in fact synonymous, but it does help us understand how Penny and Pat are thinking. Similarly, to see how things look from the perspective of Kripke’s Peter, we can work *as if* there were referentially distinct names ‘Paderewski₁’ and ‘Paderewski₂’, by allowing metaphysically impossible pseudo-worlds at which they do not co-refer, but which otherwise behave normally. Such worlds may later be epistemically ruled out for Peter by evidence he subsequently acquires. That is not a semantic insight, for the

names would in fact be synonymous (on a direct reference account, since they are actually co-referential), but it does help us understand how Peter is thinking. The models enable us to apply the formal apparatus of content-based evidential relations to such cases, in a way which takes account of agents' distorted perspectives on their own contents. These artificial modelling devices require nothing as elaborate as a fully developed framework of propositional guises, nor do they rehabilitate a genuine distinction between metaphysical and epistemic modality.

Sometimes, of course, as theorists we want to step back from such models and talk about their limitations. In that case, we need to drop the semantic fictions and separate the real contents from their various guises. We need to be able to track how the agent may have conflicting attitudes to the same content under different guises. But our ability to do so does *not* mean that guises are transparent after all. Quite generally, in describing specific ways in which a given model over-simplifies messy reality, we have to be opportunistic, using *ad hoc* means to capture specific differences. If we had a one-size-fits-all framework for perspicuously characterizing all the complexities the model ignores, there would be much less need of the model in the first place. For example, in Penny's case, the ordinary English words 'furze' and 'gorse' suffice as guises, whereas in Peter's we need to manufacture new names, 'Paderewski₁' and 'Paderewski₂'. For agents confused about the individuation of their own mental files, still further layers of differentiation may be needed, and so on. Once we recognize that transparency is an unattainable ideal, we cannot even use it as a constraint to narrow down what guises must be for that constraint to be satisfied, although under given conditions some things may do better than others in the role of guises, by permitting a closer approximation to the ideal of transparency. Unsurprisingly, on this view, real-life cognition is non-transparent all the way down.

8. Probability

Frege puzzles arise for subjective and epistemic probability too. Subjective probabilities (credences) are meant to be degrees of belief. For Fregeans, just as one can believe that Hesperus is Hesperus without believing that Hesperus is Phosphorus, that Hesperus is Hesperus can be subjectively more probable for one than that Hesperus is Phosphorus; that furze has yellow flowers is subjectively more probable for Penny than that gorse has yellow flowers. The same applies to epistemic probabilities, such as probabilities on one's evidence; they are in effect graded forms of epistemic modality. For Fregeans, one's evidence can entail that Hesperus is Hesperus without entailing that Hesperus is Phosphorus; that Hesperus is Hesperus can be more probable on one's evidence than that Hesperus is Phosphorus; that furze has yellow flowers is more probable on Penny's evidence than that gorse has yellow flowers.

For reasons analogous to those Kripke explains in 'A Puzzle about Belief', such Fregean claims are deeply problematic. Indeed, the traditional use of someone's betting behaviour to measure their credences is a paradigm of a language-sensitive heuristic for attitude ascription. Whether one accepts a bet depends on how it is specified to one. At given odds, Penny may be willing to bet that 'Furze has yellow flowers' but unwilling to bet that

‘Gorse has yellow flowers’. In this respect, the betting criterion is like the heuristics (BDP), (DP), and (SDP) for belief ascription. As is now widely agreed, betting behaviour is not definitional of credences, it is at best a fallible heuristic guide to them. For instance, someone may belong to a religious sect which considers all betting sinful; she would start accepting bets only if she left the sect, in which case her credences would be different, so her (absence of) betting behaviour does not measure her current credences. She is actually certain that betting is sinful, but if she were willing to accept bets, she would bet that betting is not sinful.

Anna Mahtani has recently highlighted ways in which Frege puzzles make trouble for current practice in welfare economics (2017, 202X). Standardly, a person’s *prospect* under a policy is defined as their expected welfare under that policy. In welfare economics, such expectations are normally calculated in terms of subjective or epistemic probabilities. One policy is *ex ante* Pareto inferior to another if and only if no one has a better prospect under the first policy than under the second and someone has a better prospect under the second policy than under the first. A reasonable principle seems to be that one should not adopt a policy if it is *ex ante* Pareto inferior to an alternative policy, which is to say that one should adopt only Pareto optimal policies. But, on a Fregean view, the prospect for a particular person under a policy can depend on how that person is presented, as Mahtani shows in realistic cases, because the subjective or epistemic probabilities used to define the prospect so depend. That seems to make the standard definition of the ‘prospect’ for the person under the policy break down, since it specifies nothing about modes of presentation. That in turn undermines the definitions of standard welfare-economic ideas such as *ex ante* Pareto optimality. If prospects are ill-defined, so is *ex ante* Pareto optimality. Mahtani proposes complex ways of defining something reminiscent of *ex ante* Pareto optimality within the Fregean framework, by generalizing across a range of admissible assignment functions from people to modes of presentation.

The problem raised by Mahtani is *not* solved by the modelling devices sketched at the end of the previous section. They merely enable one to simulate Fregean effects artificially, from an anti-Fregean starting-point: but Fregean effects are the source of Mahtani’s problem. On a resolutely anti-Fregean approach, subjective or epistemic probabilities are assigned to coarse-grained propositions, such as the proposition that x has π , where x is a person and π a property, with no mediating modes of presentation, so the original definition of expected welfare would stand, and Mahtani’s objection would not arise.

The trouble for the resolutely anti-Fregean approach is that it is so unclear how to determine the subjective or epistemic probability of a coarse-grained proposition, given the cognitive impact of modes of presentation. Indeed, the problem is even sharper for probability than for ordinary propositional attitudes. In the case of belief, we may consistently rule that one believes a proposition p if and only if one believes p under at least one guise; such a rule was tacitly followed in the discussion of Frege cases above. But if we try ruling that one’s subjective probability for p is v if and only if it is v under at least one guise, the result is inconsistency, since in Frege cases the rule will assign conflicting values to the agent’s subjective probability for p .¹⁸ The mathematical structure of probability forbids such a rule. The same problem arises for epistemic probability too. This does not mean that resolute anti-Fregeans cannot postulate subjective or epistemic probabilities at all, just that

their relation to linguistic behaviour and verbalized thought will be even more indirect than for ungraded attitudes.

However, even for Fregeans, the relation of numerical credences to linguistic behaviour and verbalized thought is already quite indirect. Without (rational) betting behaviour, it is very unclear how to determine specific numbers. The task is no easier for epistemic probabilities. They cannot simply be *read off* the linguistic behaviour and verbalized thought. Similar difficulties face anti-Fregeans if they try to assign specific numerical credences to coarse-grained propositions relative to guises.

Further challenges arise for diachronic or inter-personal applications of subjective or epistemic probabilities. Imagine a shipwrecked sailor trying to keep track of time on a desert island. When he updates his probabilities overnight, his new probability for ‘Today is my birthday’ should be his old probability for ‘Tomorrow will be my birthday’, not his old probability for ‘Today my birthday’. Similarly, imagine a young boy deferring to his mother’s opinions. His probability for ‘Today is my birthday’ should be her probability for ‘Today is your birthday’, not her probability for ‘Today is my birthday’. In such cases, letting the probabilities follow sentential guises, or Fregean senses which behave in a similar way, gives exactly the wrong results. The probabilities need to follow something more like objective states of affairs or events. To model properly what is going on, we need both yesterday’s probabilities and today’s, or both the mother’s probabilities and her son’s, to be defined over the same space of events. Individuating those events in terms of sentential guises or their Fregean analogues would send us in quite the wrong direction.

Invoking numerical subjective or epistemic probabilities *at all* is a matter of model-building, not of observation, though still justified by an appropriate level of explanatory success. That is not to say that subjective or epistemic probabilities have no basis in reality, just to acknowledge that for theoretical purposes they vastly simplify and tidy up the complicated, messy reality in order to give us some understanding of it. On that view, the simplification built into assigning subjective or epistemic probabilities to objective events may often be the best modelling choice, because it is so unclear how to do better in diachronic and inter-personal cases. The objective approach promises to be a better match with the semantics of natural language, and its capacity to make sense of theoretically powerful welfare-economic ideas such as *ex ante* Pareto optimality as they stand can be taken as another argument in its favour. Greatly complicating a good model for a small gain in descriptive detail is often a poor bargain.

Probability theory itself is a case in point. The insights it provides depend crucially on its perspicuous and tractable mathematical structure. When one tries to model extremely bounded rationality in terms of subjective or epistemic probability, it becomes very tempting to weaken the standard Kolmogorov axioms. The result tends to be only slightly closer to psychological reality, but *much* less mathematically tractable.

9. Russellianism and intensionalism

In effect, standard models of the Kolmogorov axioms for probability embody a view as to bearers of probability, including doxastic and epistemic probability. In such models,

probabilities are assigned to *events*, subsets of the set Ω of mutually exclusive possible *outcomes*. The outcomes behave just like possible worlds, and the events just like propositions as sets of possible worlds, in a Kripke model for modal logic. In effect, standard probability theory treats propositions as *intensions*.¹⁹

Intensions are not only far more coarse-grained than Fregean thoughts, but also more coarse-grained than needed for mere anti-Fregeanism. For example, on one prominent form of anti-Fregeanism, a declarative sentence in a context expresses a structured Russellian proposition, a complex built up out of the objects, properties, relations, and states of affairs to which constituents of the sentence refer in that context. Thus (26) and (27) express distinct structured Russellian propositions, since Edinburgh is a constituent of the proposition expressed by (27) but not of the proposition expressed by (26):

(26) London is in pretty.

(27) London is pretty and either Edinburgh is handsome or London is pretty.

By contrast, intensionalism implies that (26) and (27) express the same proposition, since they are necessarily equivalent. In terms of probability models, conjunction ('and') corresponds to set intersection and disjunction ('or') to set union, so (26) and (27) correspond to the same event, so there is no difference in probability. In this way, probabilistic models work like standard models of intensional semantics.

Of course, there is a temptation to deny that (26) and (27) express the same proposition, on the grounds that someone who believes the proposition that London is pretty but has never heard of Edinburgh cannot believe the proposition that London is pretty and either Edinburgh is handsome or London is pretty. However, the pre-theoretic appeal of the latter claim may itself derive from a language-sensitive heuristic for belief ascription of the kind discussed in section 6, for someone who has never heard of Edinburgh lacks normal linguistic understanding of (27), which would undermine any assent to (27) as evidence for their belief in the proposition it expresses.

Logic provides some reason to prefer intensionalism over Russellianism. For, unlike intensions, structured Russellian propositions are susceptible to versions of the Russell-Myhill paradox, a variant of Russell's paradox. The argument is roughly this. Structured Russellian propositions can do the required work in semantics only if any possible well-defined predicate determines a property. On such a plenitudinous theory of properties, there are more properties of propositions than propositions, for Cantorian reasons. But there are at least as many propositions as properties of them, for if π_1 and π_2 are distinct properties of propositions, the proposition that p has π_1 is distinct from the proposition that p has π_2 , given how Russellian structured propositions (but not intensional propositions) are individuated (p is an arbitrary fixed proposition). Putting the pieces together, there are more propositions than propositions: a contradiction. Of course, friends of Russellian structured propositions try to avoid the contradiction by restricting the comprehension principles which say what properties and propositions there are. But that makes Russellianism less natural, more *ad hoc*, than intensionalism, since the latter can consistently maintain unrestricted comprehension principles in an appropriate setting of higher-order modal logic.²⁰

Thus intensionalism about propositions has some major advantages over Russellianism. But intensionalism has an even more flattening effect than Russellianism. Perhaps the most drastic examples is mathematics. If metaphysically necessarily equivalent sentences express the same proposition, then all true sentences of mathematics express the same proposition, on the standard assumption that all truths of mathematics are metaphysically necessary. Thus Fermat's Last Theorem turns out to be the proposition that $2 + 2 = 4$. By contrast, Russellianism has no such consequence: there will be infinitely many true Russellian structured propositions about mathematics. In particular, Fermat's Last Theorem as a Russellian structured proposition is not the Russellian structured proposition that $2 + 2 = 4$; they are quite different in structure.²¹

At first sight, if intensionalism trivializes mathematics, while Russellianism does not, that is a decisive advantage for Russellianism. But Russellians should not argue that way. For the claim 'Furze is gorse' is cognitively non-trivial for Penny, even though she understands the English sentence, and it in fact expresses the same structured Russellian proposition as the trivial 'Furze is furze'. Thus Russellians should deny that the cognitive significance of a sentence comes solely from the proposition it expresses. Two sentences can express the same proposition, yet differ in cognitive significance; one sentence may be cognitively trivial, while the other is not. But if Russellians can take that line, so can intensionalists. Both sides can explain what is going on by working with attitudes to coarse-grained propositions under fine-grained guises. As seen in previous sections, there are significant limits to the explanatory value of invoking guises, but sentential guises serve well enough for the case at issue: mathematics.

Both Russellians and intensionalists can appeal to language-sensitive heuristics such as those in section 6 to explain our mistakes in ascribing attitudes to propositions expressed in the language of mathematics. Intensionalists will typically diagnose errors in attributing the *absence* of attitudes. For them, strictly speaking, anyone who knows that $2 + 2 = 4$ knows the truth expressed by a statement of Fermat's Last Theorem; attributing the absence of such knowledge is an error. By contrast, Russellians will typically diagnose errors in attributing the *presence* of attitudes. For them, strictly speaking, any difference in semantic structure makes a difference to the proposition expressed; in that respect, Russellian propositions are quite fine-grained. Thus we might describe someone who concludes $\forall x (Fx \supset Gx)$ as having concluded that every F is a G (where ' F ' and ' G ' are interpreted), but for Russellians that is an error, strictly speaking, for the formula ' $\forall x (Fx \supset Gx)$ ' and the English sentence 'Every F is a G ' differ in semantic structure. In the formula, ' F ' is a constituent of the open formula $(Fx \supset Gx)$ to which the quantifier (plus variable) $\forall x$ is then applied, whereas in the English sentence ' F ' combines with the determiner 'every' to make the complex quantifier 'every F ', which is then applied to the predicate 'is a G '. Thus the formula and the English sentence express different Russellian structured propositions.

To be more exact, *anti-contextualist* Russellians and *anti-contextualist* intensionalists will diagnose errors as just described. *Contextualist* Russellians and *contextualist* intensionalists may be able to avoid such error diagnoses. Contextualist Russellians may hold that having the attitude expressed by the attitude verb to a structured Russellian proposition which bears a contextually determined similarity relation to the structured Russellian proposition expressed by the complement sentence in the context is sufficient for the truth of

the attitude ascription in the context. Contextualist intensionalists may hold that having the attitude expressed by the attitude verb to the structured Russellian proposition expressed by the complement sentence in the context under a contextually relevant guise is necessary for the truth of the attitude ascription in the context. If granted enough degrees of freedom, both forms of contextualism may be able to avoid diagnosing systematic errors.

By either error-theoretic or contextualist means, both Russellians and intensionalists may be able to deal adequately with the apparently recalcitrant data. In that case, the overall preference should go to intensionalism over Russellianism, because intensionalism is the simpler and stronger framework.²²

10. Conclusion

My hunch is that the error-theoretic approaches will prove more explanatory than both the contextualist approaches and hybrid approaches, because error-theorists work with quite simple, general heuristics for attributing attitudes, while contextualists postulate the specific contextual restrictions on a case-by-case basis, *ad hoc*. Furthermore, we should not underestimate the significance of our pre-theoretic reactions when faced with cases like that of Pierre. Our sheer puzzlement and confusion, rightly emphasized by Kripke, cannot be explained by a smoothly functioning contextualist semantics alone. At some level, we are relying on an imperfectly reliable heuristic, whose outputs are not guaranteed to be mutually consistent. Whatever that heuristic is, we may well be relying on it in epistemology too, when we ascribe belief or knowledge in hypothetical cases. That does not warrant the panicky response that in epistemology we should stop using our ordinary means of ascribing attitudes, any more than visual illusions warrant the response that in science we should stop using vision. But it does indicate that we need some checks and balances. In particular, we should beware of letting epistemology be a largely data-driven inquiry, where so much of the data are supplied by our ordinary means of ascribing knowledge, belief, and other attitudes. We need to put more weight on theoretical virtues—such as simplicity and strength.²³

Notes

1. In this article, I will not discuss the critique of the reliability of attitude ascriptions in natural language implicit in the ‘negative programme’ of experimental philosophy. I deal with that critique at length in Williamson 2021. See also Nagel 2012.
2. Throughout this article, occurrences of terms in the complement clause of an attitude-ascribing verb are treated as semantically in its scope; if the distinction makes sense, they are *de dicto*, not *de re*.
3. For relevant discussion see Cappelen and Dever 2013 and Magidor 2015.
4. See Crimmins and Perry 1989 for an account of the envisaged kind, and Goodman and Lederman 202X for a more recent account of that kind, with extensive references to the literature. Salmon 1986 uses the apparatus of guises but his account is not of the envisaged kind. For the difference between Salmon’s guises and Fregean senses see Branquinho 1990.
5. There is no sign that the negation used in denying (4) and (6) has to be *metalinguistic*, as in ‘The food wasn’t *good*, it was *great*’. For evidence of error even in assessments of simple sentences in Frege cases, see Saul 1997 and 2010.
6. See, for example, the section on ‘Error’ in Goodman and Lederman 202X.
7. Beliefs need not be some sort of secondary meaning of ‘belief’ in English; it is just a theoretically postulated attitude to Fregean thoughts loosely modelled on belief. Presumably, there would be analogous attitudes loosely modelled on knowledge, hope, fear, and so on, though the subscript ‘s’ does not refer to an operation defined for all expressions of English.
8. Here and elsewhere, some readers may prefer to insert ‘if Hesperus exists’ after ‘that’; doing so makes no difference for present purposes.
9. Salmon 1987/88 raises some epistemologically relevant issues about Kripke’s example.
10. In semantics, epistemic necessity and epistemic possibility can be relativized to any given stock of knowledge. For Kripke, they concern specifically *a priori* knowledge.
11. For a recent book-length critique of Chalmers’ approach see Yli-Vakkuri and Hawthorne 2018.

12. For a debate on the (in)significance of the a priori / a posteriori distinction see Boghossian and Williamson 2020.
13. See Williamson 2020 for more on the philosophical relevance of heuristics to conditionals, sorites paradoxes, and semantic paradoxes.
14. Rackoczy, Bergfeld, Schwarz, and Fizke 2015 provides evidence that children acquire a basic understanding of aspectuality (guise-sensitivity in attitude ascription) at the same age, 4-5, at which they acquire a basic understanding of false belief.
15. Salmon 1986: 111 gives a similar treatment of *withholding* belief.
16. See for example Recanati 2012; the application of the idea to Frege puzzles goes back to Strawson 1971, 1974, and is made in the Fregean theory of Forbes 1990.
17. For more detailed discussion, see Goodsell 2013 and Yli-Vakkuri and Hawthorne 2018: 149-66.
18. For a proposed non-Fregean treatment of this problem see Braun 2016; for a Fregean approach see Chalmers 2011.
19. Intensions are usually defined as functions from worlds to truth-values rather than as sets of worlds, but the choice makes little difference. Ultimately, talk of propositions is best treated as a crude paraphrase of quantification into sentence position, subject to the constraint that necessary equivalence entails indiscernibility, in the setting of a higher-order logic (see Williamson 2013).
20. See Dorr 2016 for more discussion.
21. A long-standing proponent of taking intensions seriously as the objects of propositional attitudes in epistemology and the philosophy of mind is Robert Stalnaker; see Stalnaker 1984, Stalnaker 1999.
22. For the methodological implications of intensionalism for metaphysics and metaethics see Williamson 2022 and 202X respectively.
23. Written comments from Daniel Kotsi, Anna Mahtani, Jennifer Nagel, Mona Simion, and Juhani Yli-Vakkuri and conversation with Jeremy Goodman all helped to make this a better paper.

References

- Boghossian, Paul, and Williamson, Timothy. 2020: *Debating the A Priori*. Oxford: Oxford University Press.
- Branquinho, João. 1990: 'Are Salmon's guises disguised Fregean senses?', *Analysis*, 50: 19-24.
- Braun, David. 2016: 'The objects of belief and credence', *Mind*, 125: 469-497.
- Cappelen, Herman, and Dever, Josh. 2013: *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford: Oxford University Press.
- Chalmers, David. 2006: 'The foundations of two-dimensional semantics', in Manuel Garcia-Carpintero and Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*, 55-140. Oxford: Oxford University Press.
- Chalmers, David. 2011: 'Frege's puzzle and the objects of credence', *Mind*, 120: 587-635.
- Crimmins, Mark. 1992: 'I falsely believe that *p*', *Analysis*, 52: 191.
- Crimmins, Mark, and Perry, John. 1989: 'The prince and the phone booth: reporting puzzling beliefs', *Journal of Philosophy*, 86: 685-711.
- Dorr, Cian. 2016: 'To be F is to be G', *Philosophical Perspectives*, 30, 39-134.
- Forbes, Graeme. 1990: 'The indispensability of *Sinn*', *Philosophical Review*, 99: 535-563.
- Forster, Malcolm, and Sober, Elliott. 1994: 'How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions', *British Journal for the Philosophy of Science*, 45: 1-34.
- Goodman, Jeremy, and Lederman, Harvey. 202X: 'Perspectivism', *Noûs*, forthcoming.
- Goodsell, Thea. 2013: 'Mental files and their identity conditions', *Disputatio*, 5: 177-190.
- Grice, Paul. 1989: *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.
- Kaplan, David. 1989: 'Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals', in Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*, 481-563. New York: Oxford University Press.
- Kripke, Saul. 1972: 'Naming and necessity', in Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 235-355, 763-769. Dordrecht: Reidel.
- Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Blackwell.
- Kripke, Saul. 1979: 'A puzzle about belief', in Avishai Margalit (ed.), *Meaning and Use* (Dordrecht: Reidel), 239-83.
- Kripke, Saul. 1988: 'A puzzle about belief', reprint of Kripke 1979 in Nathan Salmon and Scott Soames (eds.), *Propositions and Attitudes* (Oxford: Oxford University Press, 1988), 102-248.
- Magidor, Ofra. 2015: 'The myth of the *de se*', *Philosophical Perspectives*, 29: 249-283.
- Mahtani, Anna. 2017: 'The *ex ante* Pareto Principle', *Journal of Philosophy*, 114: 303-323.
- Mahtani, Anna. 202X: 'Frege's Puzzle and the *ex ante* Pareto Principle', *Philosophical Studies*, forthcoming.
- Nagel, Jennifer. 2012. 'Intuitions and experiments: a defense of the case method in epistemology', *Philosophy and Phenomenological Research*, 85: 495-527.
- Nagel, Jennifer. 2013: 'Knowledge as a mental state', *Oxford Studies in Epistemology*, 4: 273-306.

- Rakoczy, Hannes, Bergfeld, Delia, Schwarz, Ina, and Fiske, Ella. 2015: 'Explicit theory of mind is even more unified than previously assumed: belief ascription and understanding aspectuality emerge together in development', *Child Development*, 86: 486-502.
- Recanati, François. 2012: *Mental Files*. Oxford: Oxford University Press.
- Salmon, Nathan. 1986: *Frege's Puzzle*. Cambridge, Mass.: MIT Press.
- Salmon, Nathan. 1987/88: 'How to measure the standard metre', *Proceedings of the Aristotelian Society*, 88: 193-217.
- Saul, Jennifer. 1997: 'Substitution and simple sentences', *Analysis*, 57: 102-108.
- Saul, Jennifer. 2010: *Simple Sentences, Substitution, and Intuitions*. Oxford: Oxford University Press.
- Stalnaker, Robert. 1984: *Inquiry*. Cambridge, Mass: MIT Press.
- Stalnaker, Robert. 1999: *Context and Content*. Oxford: Oxford University Press.
- Strawson, Peter. 1971: *Logico-linguistic Papers*. London: Methuen.
- Strawson, Peter. 1974: *Subject and Predicate in Logic and Grammar*. London: Methuen.
- Williamson, Timothy. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy. 2013: *Modal Logic as Metaphysics*. Oxford: Oxford University Press.
- Williamson, Timothy. 2017: 'Model-building in philosophy', in Russell Blackford and Damien Broderick (eds.), *Philosophy's Future: The Problem of Philosophical Progress*, 159-173. Oxford: Blackwell-Wiley. Reprinted in Williamson 2021a: 372-385.
- Williamson, Timothy. 2020: *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford: Oxford University Press.
- Williamson, Timothy. 2021: *The Philosophy of Philosophy*, 2nd ed. Oxford: Wiley-Blackwell.
- Williamson, Timothy. 2022: 'Metametaphysics and semantics', *Metaphilosophy*, forthcoming.
- Williamson, Timothy. 202X: 'Moral Anti-Exceptionalism', in Paul Bloomfield and David Copp (eds.), *The Oxford Handbook of Moral Realism*, to appear.
- Yli-Vakkuri, Juhani, and Hawthorne, John. 2018: *Narrow Content*. Oxford: Oxford University Press.