Introduction to Epistemology

Philipp Blum

draft version

Foreword

This course was written for an 'intensive seminar' "Introduction to Epistemology" I gave during spring term 2017 at the Faculty of Theology, University of Lucerne (April 7 and 8; May 19 and 20, 2017). I was replacing Prof. Gianfranco Soldati, from the University of Fribourg, on whose slides the course is based. I also made extensive use of my notes from the 'Kompaktseminar' "Varieties of Skepticism" Crispin Wright gave at University of Heidelberg in spring 2004,[1] as well as of my MA thesis in theoretical computer science which I submitted 2002 at the University of Berne. I learned from the feedback of my students, and from Kevin Mulligan who sent me an unpublished paper of his on correctness conditions.

To do, in this order

1. read the rest of Sellars, incorporate it into ch. 2: (Sellars 1997)
2. incorporate the rest of wright.tex - put it back into courses
3. incorporate the logic MA thesis, and the stuff on information (perhaps within ch 2)?
4. discuss Nozick and the dispositional analysis
5. make full biblio of Dretske, make proper references to zebras and closure
6. rework the scepticism paper
7. read Engisch on intentional objects and rework that chapter
8. return to correctness vs accuracy: cf Mulligan forthcoming and Buekens (on correctness)

---

[1]. What Wright said in Heidelberg about how Nozick on knowledge could be defended against counterexamples by dropping the counter-factual analysis of dispositions is identical in content (and sometimes wording) to Gundersen (2010).

Contents

Chapter 1

The Bases of Our Knowledge

1.1 Seeing Things

1.1.1 "Epistemology"

Theoretical philosophy aims to articulate a view of ourselves and the world that is systematic, comprehensive, beautiful and makes sense both of them and of our desire to understand them. Its most general articulation is within the 'subdiscipline' (if there is such a thing) of metaphysics, which asks questions about the natures and determinations of things, about the ontological categories into which they fall, their metaphysical status and their modalities.

Metaphysics, so conceived, not only concerns the world as opposed to what we think about it, but also us as parts of that world and our relations to it, themselves a part of the world, connecting two worldly items. The philosophy of mind is more narrowly concerned with such relations, in particular their foundations 'on our side', as it were – mental phenomena, such as feelings, intentions, imaginings, suppositions, beliefs, desires – and with the nature of these relations, in particular the question whether they relate things that belong to the same or to different metaphysical categories.

Among such relations between us and the world, we may distinguish two different 'directions of fit'. Some mental phenomena, such as wishes, impose conditions on the world: I determine what wish you have by asking how the world would have to be for that wish to be fulfilled. Others, such as beliefs, have conditions imposed on them by the world: asking what you believe is asking how the world *is* according to you and whether or not the world really is that way.

For mental states of both directions, we distinguish, in the determination of their content, between structural and thematic constraints. In the case of wishes, a structural constraint is that, in wishing it to be the case that p, I represent the world's being such that p as *good*. This is not a second, additional thing I do: by the very having of the wish, I represent its fulfillment as good. A thematic constraint on possible contents of wishes might be that I can only wish what is metaphysically possible: when I wish to draw a square circle, for example, I may be criticisable not just for pragmatic, moral or instrumental reasons, but also for having a wish (if it is a wish at all) that is internally defective.[1]

Epistemology, concerned with those mental phenomena that represent the world, correctly or incorrectly, as being in a certain way, and with the ties of justification among these, examines such structural and thematic constraints. With respect to belief, for example, epistemology is concerned to explain why believing that p *is* representing it as true that p, and to account for the specific type of – "epistemic" – norms and values that characterise a true

---

1. Some, e.g. Aquinas (*Sum.c.Gent.* III 25, 52, 59, 63) would even go as far as claiming that for natural wishes at least, it even has to be nomologically possible for them to be fulfilled.

belief as epistemically better than a false one and ground a specific – "epistemic" – practice of attributing praise and blame to cognitive subjects.

Translating "epistemology" as "theory of knowledge" ("théorie de la connaissance", "Erkenntnistheorie") is a theoretical choice, quite apart from the fact that "knowledge" is just one of many acceptable translations of "epistēme". While nowadays "knowledge" certainly is the key term, earlier philosophers were more concerned about wisdom, certainty or what it is rational to believe. In this introduction, however, we will follow contemporary orthodoxy and take epistemology to be centrally concerned with knowledge.

The key question of epistemology, understood as theory of knowledge, is: what can we know? It is both descriptive, aiming at giving an accurate description of what it is in virtue of which we know when we do, normative, exhibiting grounds for praise and blame in these cases and articulating a sui generis way in which intellectual enquiry is shaped by the goals that are characteristic for it, and also, thirdly, self-reflective, being a theoretical enterprise itself, thus aiming at knowledge, or truth at least.

1.1.2 Perception

Both historically and systematically, epistemology starts with perception, or at least the supposedly paradigm case of visual perception of visible properties of opaque, middle-sized, relatively immobile material things. "How do you know the stick is straight? – I have seen it, by my own eyes." It does and did not take long to realise how problematic even this paradigm case is. Here are some of the questions that may be raised.

What does the "it" refer to in the example given? What is the object of perception, what is it that we see? Many have thought it necessary to distinguish between at least the following two:

(**A**) S sees that p.
(**B**) S sees x.

The supposed contrast between (A) and (B) has been used to introduce, motivate or explicate a certain number of theoretical distinctions:

1. intensional / extensional: While we may substitute coferential terms for each other within "x" in (B) *salva veritate*,[2] we cannot do so normally within "p": if S sees Hesperus, S thereby sees Phosphorus, while seeing that Hesperus is the brightest heavenly body in the sky is not the same as seeing that Phosphorus is the brightest heavenly body in the sky.
2. propositional / objectual: While (B) relates S (at least paradigmatically) to an object / individual / particular, (A), at least on some readings, relates S to the referent of "that p", usually taken to be a proposition.
3. conceptual / non-conceptual: While propositions are normally taken to contain concepts and mastery of these may be taken to be a prerequisite for the truth of (A), a corresponding requirement seems absent for (B).
4. epistemic / non-epistemic: While (B), at least prima facie, seems direct and unmediated, (A) seems to involve some exercise of our epistemic faculties, involving something like a belief in a way (B) does not.

---

2. Unless, of course, these introduce intensional contexts of their own, as would, e.g., a substitution of "that p" for "x".

None of these uses of the contrast between (A) and (B) is unproblematic, however.[3] What is more, the binary dichotomy itself is too simplistic. In addition to (A) and (B), we also have (C), (D), (E), (F) and (G), and perhaps others as well:

(**B**) I see the stick and it is straight.
(**A**) I see that the stick is straight.
(**C**) I see the straight stick.
(**D**) I see the stick as straight.
(**E**) I see the stick as if (it were) straight.
(**F**) I see the stick being straight.
(**G**) I see the stick's being straight.

At least *prima facie*, it is not clear that all these are analysable just in terms of either the 'purely objectual' (B) and the 'purely propositional' (A). Neither is it clear how the proposed dimensions of contrast classify these cases.

'Propositional' seeing, as in (A), is not, or at least not only, a seeing of propositions. Even if propositions are perceptible, I also see the cat when I see that it is on the mat. If the proposition involved in (A) is not the primary object of the perceptual act, in what way is it involved at all? Saying that it is the content of that act is not explaining anything, but labelling the phenomenon we would like to have explained.

The objectual sense also raises further questions: do we ever see *the stick*? Perceptually given to us, it might seem, is at best its surface, or rather the parts of its surface visible from a certain point. Do we see its back-side too, the other parts of its surface, its inside? After all, sticks with different backsides would look just the same to us. If we perceive what is not given to us, how come we do not acquire information about it? If we do not, on the other hand, then how do we know, on the basis of our perception, that sticks, e.g., are three-dimensional?

As this brief and cursory survey of just some of the problems of perception reports makes clear, one of the main problem of contemporary theories of perception is to reconcile the relationality of perception, its being a way of making contact with the world, with its representational character, its being a certain take or perspective on the world.

1.1.3 The Argument from Illusion

When we see, we undergo certain physiological processes and we also have a *visual experience*. But we also have visual experiences when we are hallucinating (though, conventionally, we would not call them "seeings" or "perceptions"). Macbeth did not see or perceive a dagger before him, but instead he merely seemed to see one; he had an experience as of one or, perhaps, an experience that represented one. We can have visual experiences when we are not seeing. These are of two types:

- hallucinatory: We do not see anything, we merely seem to see something / have a visual experience as of something: there is nothing to perceive.
- illusionary: We do see something, but see it to be in some other way than it really is.

Some cases may be hard to qualify one way or other: do I see some region (wrongly) as filled when I hallucinate a pink elephant? Do I hallucinate the white wall's redness when I see it in red light? The very existence of

---

3. With respect to (1), both parts of the claim may be disputed: does Oedipus, when he sees the stranger he then kills, really see his father? does the dog, when he sees that the cat ate its food, see the cat under some mode of presentation, in a way that would, in principle (if only it could speak and reason …), allow it to describe the food-eater as a cat?
With respect to (2), we should keep in mind that the proposition that *p* may be de-se, as in "*S* sees that he himself* is making a mess" using Castañeda's 'quasi-indicator' as applied to Perry's 'messy shopper'. In these cases, the proposition does not have autonomous semantic content, and is not detachable (in the sense that "*T* saw that too" is made true by *T* (not *S*!) realising that s/he is making a mess. We also should not forget that "*x*" in (B) may refer to a fact, or another worldly item of propositional form.
With respect to (3), it may be questioned whether concept-mastery is required for all so-called 'propositional attitudes' (for example, it seems that you can regret that *p* without 'having' the concepts needed to understand that *p*); if they are not, then why require them for (A)? Conversely, it may be maintained that we need concepts to see artefacts, beginnings and endings, social events (the crash of the stockmarket) and things such as Maria's regret, the apple's being edible and the sky looking as it will soon rain.
With respect to (4), we should not forget that in persistent visual illusions, we normally continue to have the visual impression that *p* (e.g. that one of the lines is shorter than the other in the Muller-Lyer illusion) even when we not only lack the belief but even know it to be false. On the other hand, seeing *x* may be taken to be itself an epistemic relation ('acquaintance').

hallucinations and illusions may be questioned: is there ever nothing to perceive, do I ever see 'more' than how things appear to me? While I think these questions are pressing, I will follow orthodoxy in what follows and assume that illusions exist and are roughly as they are standardly described.

If it is possible to perceive things differently from how they really are or even to seem to perceive things that are not there at all, how can perception give us "immediate and direct access" to the world? Crane (2005: 4) gives the following version of the argument from illusion:

1. "When one is subject to an illusion, it seems to one that something has a quality, F, which the real ordinary object supposedly being perceived does not actually have.
2. When it seems to one that something has a quality, F, then there is something of which one is aware which does have this quality.
3. Since the real object in question is, by hypothesis, not-F, then it follows that in cases of illusion, either one is not aware of the real object after all, or if one is, one is aware of it only "indirectly" and not in the direct, unmediated way in which we normally take ourselves to be aware of objects.
4. There is no non-arbitrary way of distinguishing, from the point of view of the subject of an experience, between the phenomenology of perception and illusion […].
5. Therefore there is no reason to suppose that even in the case of genuine perception one is directly or immediately aware of ordinary objects.
6. Therefore our normal view about what perceiving is – sometimes called "naïve realism" or "direct realism" – is false. So perception cannot be what we normally think it is."

It goes without saying that all premises, as well as the inferences from (1), (2), (3), (4) to (5) and from (5) to (6) have been challenged. Acceptance of the first inference may distinguish between two (families of) views about the relationship between the phenomenology of perceptual experiences and their relationality:

Common factor theorists : If veridical experiences and hallucinatory experiences are the same in phenomenal character (what they are like to have), they must have the same type of objects (sense-datum theories and representationalism).

Disjunctivists : Two visual experiences may be phenomenally and even introspectively indiscriminable and only one of them be veridical or even have an object at all.

Another way of marking the same distinction is in terms of the direction of explanation of the 'content' of the non-veridical experience: common factor theorists take this content to be settled independently of the veridicality of the experience, while disjunctivists hold that the determination (and even the having) of the content by the non-veridical experience is irreducibly counterfactual, in terms of the content it would have if it were veridical. Common-factor theorists thus explain the good case as a complex of the bad case + whatever is needed for veridicality, while disjunctivists explain the bad case as 'botched', as something that would be a good case if only the world cooperated.

With respect to perceptual beliefs, i.e. beliefs we acquire and are justified in acquiring in virtue of undergoing some perceptual experience, we distinguish two articulations of the problem of explaining how we sometimes, but only sometimes, acquire justification for (H) on the basis of (G):

(**G**) I have the visual experience that p.
(**H**) I know that p.

Common-factor theorists seem committed to the claim that (G), because it relates me to the same object or to the same content as (A), justifies (H) in the same way as would (A). In the case of illusions and hallucinations, this is the wrong answer – how can it be avoided? The only way, it seems, is for them to deny, as in the case of (G)⊢(A), so equally in the case of (A)⊢(H), that the transmission of justification is direct: some other conditions have to obtain. But what could they be?

Sense-data theorists face this objection in the following form: if in visual experiences I am related to sense-data, then knowledge claims about ordinary mind-independent objects such as sticks can only be indirectly justified by them; the sense-data act as a 'veil', hiding the world from us. The best thing sense-data theorists can do, it seems,

to distinguish (G) from (A) is by imposing **accuracy conditions** on the sense-data: a sense-datum is *accurate* if it stands in some representational isomorphism to the scene it is a sense-datum of, i.e. pictures it in the way accurate pictures do. (G) can then be said to transmit justification to (H) iff the sense-datum in question is accurate.

Under what conditions, however, is a sense-datum accurate? Suppose I am looking at a tree and would be able, given enough time and patience, to visually discriminate exactly 18'549 leaves, each one suitably contrasting against its background. Instead of staring at the tree, I could take a high-resolution photograph and examine it in my study: it would be accurate only if it depicted exactly 18'579 leaves. When I am not counting them, however, and even though I may be said to see that the tree has more than 500 leaves and less than 100'000, it is just wrong to say that I see that the tree has 18'549 leaves. What I see is that the tree has *many* leaves, where this leaves it open how many exactly it has. The problem is that the sense-datum, whatever it is, cannot match this indeterminacy (Armstrong 1968: 220). If it is accurate, it has to depict the right number of leaves, but then it can appear to be the way the tree appears to be only if the sense-datum appears to be differently than it is. This is the answer Jackson (1977) gives to Armstrong, but it is problematic. If sense-data have a 'hidden nature' in virtue of which they can appear differently from how they are, would not then another argument from illusion (about the properties of sense-data) show that another veil has to be introduced, between us and our own sense-data, thus starting a regress?

At root the same objection can be put to representationalists as follows: It may seem an advantage of representationalism that it explains the content of the knowledge claim (H) (i.e. that it is a claim to know *that p*) on the grounds that (H) 'inherits' its content from the visual experience: the visual experience 'transmits' its content to the knowledge claim and thereby justifies it. On closer inspection, however, things are not so simple. The linguistic context for "*p*" in (H), as we have seen, is intensional; insofar as (at least some types of) knowledge require concept-possession, it is even hyperintensional: I may know that Sam is an eye-doctor without knowing that he is an ophtalmologist. Even if perception has conceptual content, however, it seems implausible to take it to be so finely grained. Suppose I recognise his profession by some external marks, in what would a perception that he is an eye-doctor differ from a perception that he is an ophtalmologist? If the content is not 'handed down' entirely from the perception to the knowledge claim, then the way it changes must be determined by the change in 'mode': because the 'modes' of (A) and (H) are different, some rules of transition have to be specified, and they will be different in the (G) to (H) and the (A) to (H) cases.

Disjunctivists also face a similar problem. While they may hold that only (A), but not (G) transmits justification to (H), they also need a story to back this up. What is it about the subjectively indiscriminable difference between cases where only (G), but not (A) is true and cases where both are true that makes (H) turn on it? The beginning of a plausible answer is to take the veridicality of perception to be 'primitive' and instead of explaining (A) as '(H)-plus', explain (H) as '(A)-minus'. One way of doing this is in terms of **success conditions**: a perceptual experience is successful iff it is a perception, and only successful perceptual experiences ground knowledge claims. Requiring truth for success, however, is characteristic not just of perception, but also of belief and, to some extent to be determined later, of emotions as well. The central explanandum is thus the relation that is required to hold between the state reported by (A) (but not every state reportable by (H)) and the state that is claimed to obtain by (G): a relation of grounding and justification I will call the "basing relation" in what follows.

1.1.4 The Intentionality of Perception

Let us consider one version of the argument from illusion:

  1 A stick in front of me looks bent.
  2 A stick in front of you looks straight.
  3 There is at most one stick in front of us.
  4 No stick can be both bent and straight.
∴ 5 In at least one case, one of our perceptual experiences is non-veridical.
  6 In the non-veridical perceptual experience, one is presented with a mind-dependent intentional object.
  7 Both perceptual experiences have the same type of intentional object.
∴ 8 The intentional object of both experiences is mind-dependent.
  9 Sticks are not mind-dependent.

∴ 10 I neither case do we see the stick.

Hence nothing that may look to have contrary properties is ever seen – a *reductio*. While most attention has been focussed on denying (6) or (7) or both, I would like to focus attention on the description of the situation, i.e. on (1) and (2). My main motivation is that already (5) is implausible: while I may certainly form false beliefs on the basis of one of my perceptual experiences, they themselves are not necessarily misleading: at least on one sense of "looks", the straight stick half-immersed in water does look straight, i.e. looks as straight sticks half in water do.

Austin denies that (1) and (2) are compossible. Irrealists deny (9), and may therefore also deny (3). (6) appears vulnerable, but depending on how we understand 'intentional object' and 'mind-dependent', it allows of uncontroversial interpretations. (7) is denied by disjunctivists. The phenomenal principle, or something like it, may be appealed to justify (5), by deriving it from (1), (2), (3) and (4):

  1   A stick in front of me looks bent.
∴, 11   There is something of which I am aware that is bent.
  2   A stick in front of me looks straight.
∴, 12   There is something of which I am aware that is straight.
  3   There is at most one stick in front of me.
  4   No stick can be both bent and straight.
  13   There is a most one thing of which I am aware in the two cases.
∴, 14   Neither in (1) nor in (2) am I aware of the stick.

This is not a particularly good argument, however: the conclusion is too strong (as it incorporates already a version of (7)), and the additional premise (11) is hard to justify. Johnston, for example, denies (5) because he denies (13) – there are two sensible profiles of which I am aware, one of which is a proper part of the other. How does Johnston, however, account for (1) and (2)?

The most promising way out, I think, is to block the deduction of (5), while preserving a sense in which both (1) and (2) may be true. A first option is to attack (4). One may hold that the properties in question (being bent, being straight) are not (parts of) the intentional object of these perceptions, but rather of their contents, conceived of as modes of presentations of the object: the stick is presented in the straight-mode or in the bent-mode. Alternatively, one may hold that being presented in the straight- (or bent-)mode are not representational at all, but rather sensational properties, qualia of the experiences themselves. The problem with these views is that do not address the real problem: even if we grant that bentness may be a way the stick is presented to me or a qualitative property of my experience, it is certainly *also* a property of the stick: in one of the two cases at least, the roundness present in my experience is 'matched' by a roundness of the stick – and we have to find a ground for saying that this is not the case for bentness.

Adverbialists reinterpret (1) and (2) as

  1$^\text{i}$ I am appeared to bent-ly.
  2$^\text{i}$ I am appeared to straight-ly.

Even though we do not then have analogues of (3) and (4), (5) *still* seems to follow: If we are appeared to bent-ly and straight-ly, then in at least one case, the perceptual experience is non-veridical. We also have trouble explaining why (4), or at least some version of it, belongs to a exhaustive description of the situation: to see this, compare the scenario with what changes when we take the stick out of the water: even if we keep our respective perspectives on the stick, an important source of conflict and disagreement disappears. This source, whatever it is, is not captured by (1$^\text{i}$) and (2$^\text{i}$). In this second, but not the first respect, representational views fare better:

  1$^\text{ii}$ I see the stick as bent.
  2$^\text{ii}$ I see the stick as straight.

Insofar as bentness and straightness may figure in contents (as represented properties, rather than as being exemplified by things as sticks), we do account for the conflict noted in (4). Given (3) and (4), however, it still follows that in at least one case one of us sees the stick as F where the stick is not F – so (5) still follows.

The sense data theorist reinterprets (1) and (2) as

1^{iii} It appears to me: something is bent*.

2^{iii} It appears to me: something is straight*.

Bentness* and straightness* are not properties of sticks, but of sense-data or of experiences: if they can both be had by the same sense-datum, then we do not have an analogue of (4); if they cannot, then we do not have an analogue of (3). The victory is Pyrrhonic, however, because even if they can perhaps provide a reading of "seeing" on which it is true that I see the stick, I will not on their view see its bentness or straightness.

Perspectival realists also deny (5) by reinterpreting (1) and (2). Their version is:

1^{iv} The stick appears-bent-to-me.

2^{iv} The stick appears-straight-to-me.

This, I think is part of the right response to the argument. It combines the adverbial account of the (allegedly) contrary properties with the introduction of appearance properties – properties like *being bent-to-me*, or rather: *being-bent-from-here*, the relativisation encapsulating the information about the position and perceptual apparatus of the observer and the information about the perceptual *milieu* we would need to appeal to correctly predict from what other actual or counterfactual experiences the perceptual experience in question would be subjectively indiscriminable.

Bent-to-me and *straight-to-me* are not incompatible, so we do not have an analogue of (4); but neither are they properties of experiences, nor of sense-data, nor of points of views or observers. Are they properties of sticks? In a sense yes, but also in another sense no. The property the half-immersed stick appears to have from here, i.e. the one I can correctly describe by painting a inflected line, is a property the stick no longer appears to have once I have taken it out of the water. It is very well possible, however, that the half-immersed stick did not to you appear to have this property, e.g. if you were looking at it from another angle it may have looked straight to you, even when half-immersed in water. When we took the stick out, it started to appear differently to me, but not to you. While *bent to me* certainly was a property the stick appeared (to me) to have, it was not so for you and is now not so for me. When we immersed it in water, there was a look the stick acquired, a look it did not have before: by changing the perceptual milieu and the set of available perspectives on the stick, the stick acquired a new look it did not have before, a look I describe ascribing the perspectival property *bent-to-you* and you describe ascribing the different, but not incompatible property *straight-to-me*.

Objects of perception are things from a certain point of view, i.e. perspectival facts or 'looks'. The points of view accounting for their perspectivality are not the perceptual acts, nor are they modes of presentation of the object (if there is one): they are located on the 'object side' as it were. Take, to change the example, the round coin lying on the table in front of me, looking, at we would say, oval from here. The same coin, it seems, can look round from certain points of view and elliptical from others. As before, *round a from certain point of view* and *elliptical from another point of view* are not incompatible properties, so the coin does not look to have incompatible properties. The cup always appears round from here or there. To appear round from here means that the cup appears to be at a certain *distance* and *orientation* from the given point of view. Orientation is reducible to relative distance of parts: the orientation of the cup changes from a point of view to another iff the ratio of distances between each part and the point of view changes. For instance, from some points of view (right above of the cup), all the parts of the edge are at equal distance from the point of view; while for some other points of view (from the side of the cup), there are parts of the edge that are closer to the point of view than other ones. These ratios, however, should not be confused with the ratios among the parts of the edge themselves: they are all at the same distance of each other, and appear to be so in the perceptual experience. Ellipses then are just artifacts from painters that project three-dimensional oriented objects on two dimensional vertically oriented screens. When we adopt the attitude of the painter, we create an illusion from a perception: we change the orientation and shape of the perceptual objects. We change the relative distances of the edges of the cup to the point of view, thereby creating the illusion of having an elliptical object in view.

These perspectival appearance properties are thus not features of perspective-independent things, but rather of appearances, appearances *of* perspective-independent things, to be sure, but not things that are themselves independent of perspectives. Kant is right in thinking that the perception of appearance properties is the perception of

appearances, where such appearances are grounded in the things they are appearances of. It is the appearance of the stick half-immersed in water that looks bent-from-here and the same appearance that may look straight-from-here with respect to some other location or observer. It is the change in appearance that is responsible for the fact that once the stick is taken out of the water and seen by both of us in clear air, it may no longer be said to appear bent in any sense. Because appearances are grounded in the things appearing, together with their environment, they are appearances of things – things with which we are in perceptual contact when we perceive the appearances. Can we thus say that the things are only indirectly seen, through their appearances as it were?

We should not, however, in the way of sense-data theorists, take these appearances to be the direct objects of perception – rather, it is their grounds that are seen through them. This explanation of appearance properties thus takes us to

1^V The stick-as-seen-by-us appears-bent-to-me.

2^V The stick-as-seen-by-us appears-straight-to-you.

We thus have two perceptions, but still just one stick appearing to both of us. (5) does not follow if the two appearances really have their appearance properties. Such a reply to the argument from illusion, I surmise, is not only independently plausible, but also generalises to the analoguous arguments by McTaggart against temporal A-determinations and and by Frank Jackson in favour of perceptual qualia.

Perspectival facts are not sense data, but are 'out there', full citizens of the mind-independent external world. They are not parts or aspects of the experiencing subject and they exist independently of experiencing subjects. They are perspectival only in that they contain a perspectivally 'modulated' property, 'modulated' by its being relational with respect to a point of view. A point of view is just a point in space (compare Atkin (2006)), where a (point-sized) seeing eye can be located, but that can as well remain unoccupied. Actual points of view do not imply actual views. Points of view are not acts of perception, nor part of the act of perception: they do not depend on perceiving objects. Searle is thus wrong to assert that aspectual shape "cannot be exhaustively or completely characterized solely in terms of third person […] predicates" (Searle 1992). Rather, we advocate a return to some robust form of realism about perspectival facts, like the one that was endorsed by neo-realists such as Nunn (1910) or Holt (1912).

Before drawing such morals, however, we have to know more about sticks-as-seen-by-us, i.e. appearances of sticks.

When I say that the wine tastes sweet, I am qualifying the taste as sweet, and only secondarily the wine as having a sweet taste.[4] On one reading, the first qualification is not subject to doubt: in this sense, I am qualifying the taste as tasting-sweet-to-me, something which you cannot dispute. Sometimes, however, I make a stronger claim, not only that the taste is sweet-to-me, but that it is sweet *tout court*, and so that you should taste it as sweet as well, if you are tasting properly. Here we have conceptual space for taste courses and experts, judges and faultless disagreement.

Another dimension of possible disagreement opens up with the secondary qualification: I may be wrong that it is the wine that tastes sweet. This suggests that the 'extension' of my assertion is via a *causal* link: I am saying of the taste that it is sweet (and there cannot be doubt about this) and of the wine that it has that taste, i.e. causally produced it (and about this I could be wrong).

Things stand differently with colours, or rather with chromatic profiles:[5]

> "What puts vision apart from hearing, smell, and taste is that we do not conceive of the visible world as offering us objects of visual awareness and attention distinct from (but coincident with) the concrete objects that we also see." (Martin 2010: 187), citing O'Shaughnessy (2000: 571–572)

The chromatic profile of a coloured thing (its visual appearance, its look) is more closely tied to it than its taste or its visual shape: the being-the-appearance-of relation here is not causal, but constitutive. Looks are nothing but looks of things, i.e. things looking a certain way. In other terms: for a to have a visual look that appears red is both, for the look to be red and for the thing itself to appear red. Two possible sources of error present in the case of

---

4. The "primary"/"secondary" contrast is from Martin (2010: 183).

5. The switch to chromatic profiles is needed to preserve Aristotle's claim that they are the proper sensibles of sight: we also see black and white things, transparent things and things that do not have surfaces, such as shadows, holes, holograms and rainbows.

taste are thereby eliminated: it is not possible that only its look, but not the thing looking a certain way appears red, and the look itself cannot just appear red but fail to be red.

What Mary gets newly acquainted with, when she gets out of her black-and-white room, are not new properties, but new things: the looks things have when they appear to you e.g. in clear daylight. Being an eminent colour scientist, she knew everything there is to know about colours; being equally an eminent psychologist of colour perception, she also knew everything there is to know about visual appearance properties, properties such as *looking red from here*. What she did not know, however, are an important class of things, visual looks – ie. the things that primarily have the colours and in virtue of which coloured things appear the way they do. Instead of

1^{v} The patch-in-this-light looks-red-to-me.
2^{v} The patch-in-this-light looks-green-to-you.

we thus have the simpler:

1^{v} The patch-in-this-light is red.
2^{v} The patch-in-this-light is green.

H is a perfect hallucination and has a phenomenal character and H is subjectively indiscriminable from a perception P, but differing from P in phenomenal character (as per relationalism). You do not have to deny infallibility of introspection in the favourable circumstances (and deny what Locatelli calls 'superficiality'), ie you do not have to claim

1 how H is $\neq$ how H introspectively seems

for we are in the presence of:

2 how H introspectively seems $=$ how H introspectively is
3 how H is $=$ how H introspectively is

(2) and (3) can be granted, ie (4) is fine

4 how H is $=$ how H introspectively seems

But what we do not need to grant is that this means that H cannot differ in phenomenal character from a subjectively indiscriminable perception. So we have that (5) and (6) are compatible in the presence of (7):

5 how H is $=$ how H introspectively seems
6 H and P cannot be distinguished introspectively but have different phenomenal characters.
6a how P introspectively seems $=$ how H introspectively seems
6b how P is $\neq$ how H is
7 how P is $\neq$ how P introspectively seems

So what P is subjectively indiscriminable from does not exhaust its phenomenal character. This is compatible with introspection being infallible with respect to P too:

8 how P introspectively is $=$ how P introspectively seems

but it requires that it has a broader nature than what is available through introspection (eg, to what hallucinations it is introspectively indiscriminable from).

Appearances more generally

To speak of "*x* as it appears" or "*x* as represented" is ambiguous in at least three ways:

process the processes of appearing and of representing are different – they differ in their direction, and in what they are grounded in: things appear in virtue of how they are, while subjects represent in virtue of how they are and these grounds are different in the case of the representation of mind-independent matters of fact;
event the two processes may still coincide, and thus be the same event: every appearing is then also a being represented, and every representing also a being appeared to: it is in virtue of such coincidence that these events make available the same information, and reveal the same aspect of the world;

result the one event that is both adequately described as "*x* appears to *y* as *F*" and "*x* is represented as *F* by *y*" may still have two different results, and modify its two relata in two different ways: as a dancing 'produces' both a dancer and a dance, an event of representing/appearing produces both an appearance and a representation.

1.2 Intentionality and Representation

1.2.1 Emotions and their Correctness Conditions

Success- or, more generally, **correctness conditions** explain why emotions such as fear represent the world without being propositional attitudes. Emotions, at least in standard cases,[6] stand 'between' perceptions and judgements with respect how 'explicitly' representational they are and thus afford a model to understand how something may have a mind-to-world direction of fit without being a description of the world that is either true or false. Contrary to some, we thus distinguish emotions both from axiological judgements ("This is dangerous") and from axiological perceptions (seeing the dog as dangerous).

Suppose I am standing in a Roman amphitheatre, about to be eaten by a lion. Looking at the hungry, angry, ferocious lion before me, I note certain characteristics, of me, the lion and the situation, that induce me to believe that the lion is dangerous (to me, in this situation) and also to see it as dangerous. In feeling fear of the lion, the world, and some aspect of it in particular, is presented to me as dangerous. By feeling the fear, not just as a consequence of it, I am motivated to flee – this marks a contrast to the judgement that the lion is dangerous. By feeling the fear, not just as a consequence of it, I am in contact with danger and in a position to acquire the concept of danger – this marks a contrast to the perception of the lion as dangerous.

Taking, with Aquinas, danger to be the formal object of fear, explains this double character of emotions: that they side with perceptions in being themselves motivational, but with judgements in making demands on the world that go beyond how things appear to us to be. That the formal object of fear, danger, is something I evaluate negatively in addition explains why fear is a *negative* emotion, an emotion I would rather not have. That I would rather not have it is a consequence of my metaphysically and logically prior desire to live in a world that is safe (for me); given, however, that I live in a dangerous world, fear is often the appropriate emotion to have.[7]

Emotions share with perceptual experiences that they may persist even if they are known to be incorrect: even if I know that my fear of the lion is based on a hallucination, I still feel it, in the same way I still 'see' the lines in the Muller-Lyer illusion as of different length even when I have measured them and found out not to be. They differ from perceptions, however, in that they may be incorrect in two different ways, by being based on an illusionary or hallucinatory perception (the lion is in a cage or there is no lion at all), but also by leading us astray about the axiological properties of their direct object, representing a harmless cat or a robot as dangerous. In this latter case, I not only, as in an illusionary or hallucinatory perceptual experience, have an inaccurate representation of the lion, but I also have the wrong *reaction* to it – whereas, at least in cases of persistent and not self-induced illusions, there is nothing wrong with what I do as a perceiver when I 'see' the stick half-immersed in water as bent (how else should I see it?). In some cases, moreover, emotions can be perfectly appropriate even *in the absence* of any direct object, as with *Angst* or the fear of a forthcoming exam (which is itself present, not just an anticipation of some future episode of fear). In such cases at least, whether or not an emotion is appropriate not just as a reaction to the belief but as a reaction to (axiological aspects of) the world depends on the presence or absence of its formal object: it is in this sense that the formal object determines the correctness conditions of the emotion.[8]

Like perceptions, emotions provide prima facie and pro tanto reasons for axiological beliefs, whether or not they are themselves justified by those beliefs being true.[9] The axiological beliefs are *intentionally* about the formal object:

---

6. I am setting aside here other, non-standard, types of emotional episodes, such as they occur in cases of cinema emotions, vicarious emotions (such as those, perhaps, involved in empathy), emotional contagion etc.

7. I understand this last claim contrastively: given that I react emotionally to the situation at all, fear is one of the right such reactions. We will return to "appropriate" below.

8. I leave open the additional claim that emotions are *individuated* by their formal objects rather than, e.g., by how they feel or by how it is to have them.

9. Whether this second, additional claim is also true depends on whether 'buckpassing', 'sentimentalist', 'recessive' or 'fitting' accounts of value are correct, according to which the lion is dangerous, in part at least, because it is appropriate (to me, or some normal person) to fear it.

they represent it as being present even when it is not, e.g. when I inappropriate fear the harmless house-spider in front of me. Insofar as the belief inherits this directedness from the emotion that bases it, the emotion itself exhibits such intentionality, or 'presence-in-absence' of the formal object.[10]

It is this intentional presence of the formal object that explains why it is irrational to continue to have 'recalcitrant' emotions, i.e. emotions known to be inappropriate. It is irrational because the combination of my knowledge and of my emotional state commit me to taking the world in a way it cannot be. We have here a *sui generis* incompatibility of different attitudes, which is not an incompatibility of their contents; the axiological properties that justify my emotional reaction and to which, as their intentional objects, my emotions give me access to are not represented by them, nor by their content. Even though not represented, such properties are made available to beliefs that are not just intentionally, but referentially about them (Mulligan 1998: 172). That (J) and (K) depend in the same way on (I) can thus explain why (J) justifies (K):

(**I**) The lion is dangerous.
(**J**) I fear the lion.
(**K**) I believe that the lion is dangerous.

Because (J), for its correctness, depends on (I), basing (K) on (J) allows for the transmission of warrant: the justification I have for (J) may be credited to (K), and may account for my knowledge that the lion is dangerous.

Emotions, in this way, are intentional (directed towards some formal object, even when the formal object is absent and the emotion inappropriate), *about* an axiological feature of the world, without thereby ascribing or predicating of some worldly item that it exemplifies the feature in question. To understand better this difference between correctness and truth conditions (the first of which emotions have, the second they lack), we have to examine further the metaphysical differences between intentionality and representation.

two objections to the theory of correctness conditions: correctness conditions are not truth-evaluable and one central type of correctness condition, for judgement and belief, is superfluous.

correctness conditions for psychological states that exhibit intentionality (from Mulligan (2007: 207)):

x desires to Fx	x ought to F	"Tun-sollen"
x wishes that p	It ought to be the case that p	"Sein-sollen"
x values y	y is valuable	
x admires y	y is admirable	
x fears that y	y is dangerous	
x 'values' that p	That p is valuable, is a "Wertverhalt"	
x regrets that p	It is regrettable that p	
x is ashamed that p	It is shameful that p	
x prefers y to z	y is better than z	
x judges (believes) that p	The state of affairs that p obtains	
	The proposition that p is true	
x conjectures that p	It is probable that p	
x has an interrogative attitude toward p	It is questionable whether p	
x doubts whether p	It is doubtful whether p	
x is certain that p	It is certain that p	

Mulligan (2007: 208): it includes an explanatory claim: If x correctly judges that p, then (x correctly judges that p because the state of affairs that p obtains). If x correctly judges that p, then (x correctly judges that p because the proposition that p is true). If x correctly conjectures that p, then (x correctly conjectures that p be- cause it is probable that p).

In holding that sentimentalists and buckpassers can maintain both that I (correctly) fear the dog because it is dangerous and that it is dangerous because I (correctly) fear it, as long as they distinguish between the epistemic 'because' in the first, and the 'metaphysical' 'because' in the second claim, I depart from e.g. Brady (2010: 122–123).

10. I am following here the now orthodox conception of formal objects of emotions (cf. Teroni 2007), in contrast to e.g. Kenny (2003: 132) who takes the formal object of ϕing to be "the object under that description which *must* apply to it if it is to be possible to ϕ it".

Mulligan (2007: 209-210): while mental states and acts or their contents represent (conceptually) their satisfaction conditions, at least some of them do not represent their correctness conditions. To undergo an emotion is indeed to stand in an intentional relation to value, but the relation is not belief nor does it involve any representation (thought) of value.

1.2.2 Two Types of Perspectivality

What we see, what we feel and what we represent have their perspectivality in common. This does not by itself entail that their perspectivality has to be explained in the same way, nor even that it is of the same type. The perspectivality of perception is partly physiological, partly a question about the perceptual milieu and partly a matter of the relative orientation of perceiver and perceived: what I see when I see the stick is how half-immersed sticks, in this situation, look to me from where I am. The perspectivality of emotions is at least in part explained by their being reactions by specific emoters to specific situations: what is dangerous to me is perhaps not dangerous to you. This may justify an account of the formal objects as being itself perspectival: the correctness of my fear could then be said to turn not on danger *tout court*, but on danger-for-me-now-here. With representation generally, many different factors may matter for perspectivality: while some of them may also be attributed to the formal object (e.g. assertability-in-the-philosophy-seminar), or even to the direct object (e.g. Hesperus rather than Phosphorus), others are types of medium- or message-specificity (e.g. 'analog' representations exploiting my contingent discrimination thresholds, de-se or de-nunc representation that makes the content itself dependent on the representer or, even more specifically, the act of representing itself).

I hope some clarification may be achieved by distinguishing two closely related, but distinct features of at least some mental states: intentionality and representation. Here are their, rough and provisional, characterisations:

Intentionality A mental state is intentional iff it is 'directed' towards something outside itself, where this directedness exhibits the 'presence in absence' feature, i.e. a 'pointer' 'standing in' for something (at least potentially) absent.

Representation A mental state is representational iff it is either correct or incorrect, i.e. can be assessed in terms of its 'fit' to the world, i.e. in terms of what (if any) information about the world its occurrence has.

According to Brentano, intentionality is the mark of the mental: it is the feature whereby some internal state reaches out to the world, directing the mental subject towards such (potentially absent) features. Correctness conditions specify such intentional content, but – being conditions – do not themselves require this content to be satisfied. If I am looking for the Holy Grail, for example, my activity is directed towards, and rationalisable only with respect to the Holy Grail, which, or so we shall assume, does not exist. I am intentionally directed towards the Holy Grail, without standing in a relation to it: there is nothing, after all, for me to stand in a relation to. While searching, it is in my search that the Holy Grail is 'present' to me, as the thing I am looking for; that it is 'present' to me here just means that there are conditions, if it is non-pathological, that determine when my search would be successful. For this, I do not have to be able to describe the Holy Grail, and even less to uniquely identify it; it is enough if 'I know it when I see it' and would then regard my search as successful.

According to Dretske, it is transmission of information, in particular along causal links, that explains how things like us succeed in representing the world. While natural signs, such as smoke indicating fire, may be (and, according to the orthodox Shannon-Weaver theory of information, are) information about their causes, they do not encode this information in a format suitable for its transmission. Information transmission in full-blown human communication, on the other hand, is incredibly complicated: not only do people lie and mislead, neither mean what they say nor say what they mean, but a number of complicated 'uptake' conditions on the side of the hearer must be fulfilled for information to be transmitted (communicative acts have to be understood, in some sense of this slippery term). Dretske's paradigm cases lie in between these two extremes: measurement devices such as speedometers are not 'natural' signs for what they represent, even though they stand in causal contact to it; given their adequate design and proper functioning, however, they transmit information independently of being read or understood in any way.

Both intentionality and representation 'involve', in some way, the world, or something outside their media; both give us a sense in which mental states are 'about' something else. They do so, however, in slightly different ways: they differ in how they cross-cut two distinctions commonly identified with each other: while representation is intrinsic, but also relational, intentionality is non-relational, but also extrinsic.

As commonly introduced, intrinsic properties of *a*

1. are / account for / ground 'how *a* is by itself', are exemplified by *a* 'in virtue of the way it is in itself';
2. make for genuine similarity, are 'non-disjunctive', have 'non-gerrymandered' extensions;
3. are shared by *a* and its duplicates / replicas / perfect copies.

It is is the first feature that excludes *having a brother*, for that property of me involves my brother; it is the second that excludes *grue*, for being grue is being green and examined before *t* or being blue and examine at or after *t*; it is the third that excludes *not being accompanied by a unicorn*, for that property could be lacked by a perfect replica of the entire universe, existing in a larger world that also contains some unicorn.

In some metaphysical systems, intrinsic properties are supposed to play certain theoretical roles. They

1. are qualitative natures of combinatorial units;
2. make for real, as opposed to Cambridge change;
3. do not entail, nor are entailed by the existence of any other things wholly distinct from their bearers.

Combinatorial units – "substances" in one sense of this term – are the elements that are recombined when describing alternative possibilities. Socrates becoming taller than Simmias is a real change for Socrates when Socrates grows, but a Cambridge change for him if Simmias becomes smaller. A (neo-)Humean ban on necessary connections between distinct existences is restricted to their intrinsic properties: that my being such that Socrates is white entails that he is white is not a reason to deny me that property, but rather a reason to think it is not non-relational.

It has turned out surprisingly difficult to turn this intuitive notion into a precise definition. According to the most discussed proposal, by Lewis & Langton (1998), a property is intrinsic if it does not distinguish between things that have the same pure, non-disjunctive and non-co-disjunctive properties that are independent of loneliness and of accompaniment.[11]

At least prima facie, the relational/non-relational contrast is different. Non-relational properties of *a*

1. do not 'essentially mention' other things than *a*;
2. do not 'stem from' metaphysically / conceptually / explanatorily prior relations *a* has;
3. are 'genuinely monadic';

Saying of me that I am Michael's brother 'essentially mentions' Michael – which is why it is an ascription of an 'impure' property. It is also a derivative property, presupposing (and mentioning) the prior relation of brotherhood that obtains between me and him, and it is for this reason not 'genuinely monadic', but a de-relativisation of a relation, i.e. a relational property.

Relational properties are also supposed to play certain theoretical roles. Non-relational properties

1. are wholly qualitative: their nature is exhausted by how the things that have them are;
2. are non-haecceitistic: may be shared by distinct indiscernibles;
3. are pure, i.e. entirely general, i.e. do not involve particulars.

If Michael has an indiscernible twin, everything qualitative ascribable to me by predicating "…is Michael's brother" could still be true of me, and I could still lack the property (being rather the brother of his twin); conversely, my qualitatively identical twin could lack it if Michael does not coexist with him; hence, a general description of the

---

11. A property is pure iff its exemplification does not imply the existence of anything else than the thing exemplifying it. Something is accompanied iff it does not coexist with a contingent wholly distinct thing and it is lonely iff it coexists only with its proper parts (if it has any). A property is independent of loneliness (accompaniment) iff it is both possible that is is had and that it is lacked by a lonely (accompanied) thing. A property is *disjunctive* iff it can be expressed by a disjunctive predicate but is not natural and much less natural than either of its disjuncts. The pure, non-disjunctive and non-co-disjunctive properties independent of loneliness and accompaniment are called "basic intrinsic" by Lewis and Langton. Their definition says that a property is intrinsic iff it supervenes on basic intrinsic properties, or, equivalently, iff it never differs between duplicates (where two things are duplicates iff they have the same basic intrinsic properties).

world, not involving names or referential devices, will not fix whether or not I am *Michael's* brother (rather than the brother of his indiscernible twin).

Attempts at defining relationality of properties have mainly focussed on purity. A property *P* is called "impure" iff there is a relation *R* and a *y*, such that whenever anything, *x*, has the property, it also stands in relation *R* to *y*. Metaphysically, relational properties have been characterised as properties that are individuated with reference to relations (Hochberg 1988: 196): to say that, generally and as a matter of logical truth, if $a = b$, then $\lambda x(aRx) = \lambda x(bRx)$, we need to quantify over relations.[12]

Though this is not widely remarked in the literature, it seems to me that so-characterised at least, the two distinctions crosscut. Here are some examples of the relational intrinsic:

- *having a as a part*: this is intrinsic because it only turns on how its bearer is by itself, but relational, because it mentions *a* (and not it's duplicate!) as its part.
- the value of Diana's dress: this is its intrinsic value because it is not determined in terms of what you can buy for it, why you want to have it or any other external determinants, but is the value it has in virtue of what it is in itself, i.e. in virtue of being *Diana's* dress; it is still relational, however, because a dress worn by her indiscernible twin would be (much?) less valuable.
- *being of a crime* of some punishment: this is intrinsic if the punishment is reserved for this crime, and intimately depending on it, as e.g. a specifically destined act of reparation is; it is still relational, however, for the punishment relates the punished to the crime they committed: qualitatively the same punishment for another crime would relate the punished to something else.

Here are some examples of the non-relational extrinsic:

- *not being accompanied by a unicorn*: this is an extrinsic property of everything there is because everything there is could be just as it is and a unicorn exist in addition; it is not a relational property, however, because no relation can relate you to something that does not exist.
- *being all there is*: this is extrinsic by the same token, but also non-relational, unless you posit ontologically dubious and (arguably) paradoxical existing 'totality states of affairs'.
- *being surprising* of an event: this is extrinsic, because an event has it only if embedded in a certain context, within which it is surprising; it is non-relational, however, because you can wholly and completely appreciate this characteristic without remembering or otherwise knowing what has gone before.

1.2.3 Intrinsicness and Relationality

Intentional properties are extrinsic, but non-relational

Zwei gegensätzliche Tendenzen:

- Emotionen als axiologische Urteile. Vorteil: erklärt den repräsentationalen Charakter; Nachteil: lässt die motivationale Kraft mysteriös erscheinen.
- Emotionen als axiologische Wahrnehmungen. Vorteil: erklärt die Rechtfertigbarkeit von Urteilen und vielleicht die Motivation von Handlungen; Nachteil: mis-charakterisiert den repräsentationalen Charakter (Illusionen).

General problems: - how to bridge the gap from the axiological to the deontic and then to the motivational - duality in direction of fit: represents the world, but as negative or positive (IS THIS CAPTURED BY THE NOTION OF VALENCE) – cf Aquinas, emotions (desire and disgust) are characterised by a movement-towards and a movement-away-from - why think that emotions are individuated by correctness conditions (and not, eg., by how they feel)?

Möglicher Mittelweg: eine Emotion ist korrekt, wenn ihr materiales Objekt ihrem formalen Objekt entspricht. Vorteile: erklärt die Begründungsfunktion, verlangt keine Wahrheit, lässt Supervenienz zu, schliesst möglicher-

---

12. The reason why loving-Superman and loving-Clark-Kent is one and the same property (and Lois Lane, as a matter of logic, exemplifies one iff she exemplifies the other), is that Superman is Clark Kent; therefore, the properties are not atomic, but derelativisations of the prior relation of loving.

weise eine Brücke von 'axiologisch' zu 'deontisch'. – ENTSPRICHT: SUPERVENIENCE, OR ANOTHER NOTION THAT PERHAPS ALREADY PRESUPPOSES CORRECTNESS CONDITIONS?

According to what Chisholm (1952: 56) calls "Brentano's Thesis" – that intentionality is the mark of the mental – intentional properties are extrinsic, but non-relational. They are extrinsic, because they are signs, but non-relational, because they are characterised by "intentional inexistence": psychological states may exist even in the absence of what they are about.

The representationality of some properties has to be sharply distinguished from their intentionality. A property of something is intentional iff it is *taken to be* about something else than itself. It is so taken to be if we attribute to it conditions under which it may be said to be correct. Correctness conditions specify the intentional content, but – being conditions – do not themselves require this content to be satisfied.[13]

Because they are outward-directed, and cannot be accounted for without reference to their intentional objects, intentional states are extrinsic: they are what they are in virtue of participating in a complex process, which not only involves their objects and their bearer, but also a process of interpretation or understanding.[14]

Intentionality is the flip-side of representation: whereas representational properties are intrinsic, but relational, intentional properties are extrinsic, but non-relational. Taking up an attitude towards the cognitive base turns the latter's relatum into the former's 'intentional object'. Such a 'conversion' of the relatum of a representational state into the intentional object on which an intentional state depends for its existence without being related to it, is what happens in Kantian 'synthesis': when I see a thing as red and white, redness and whiteness hang together by being aspects of the one thing my perception relates me to; when I, however, only imagine a red and white thing, the link can not come from the object alone – it must be 'constructed' by my faculty of imagination, and it is so constructed by my imagining *one* thing as both red and white. This intentional object will therefore be extrinsic, depending for its existence on my act of taking my representational state in a certain way.

Representational properties are intrinsic, but relational

Virtually any state can be used as a representation: you can decide to use a red flag to represent danger and you can take your aching muscles as a sign that you should not have walked that far. Such representational contents, however, derive from contingent dispositions to interpret the relevant states in certain ways. Since you could equally well take red flags or aching muscles to represent something different (that the communists are marching or that your work-out has been successful, for example), neither the red flag nor the aching muscles have their representational content intrinsically. Independently of what properties they intrinsically have, we *use* them as signs for other things, we *bestow* on them the representational powers they have. Not all powers of representation, however, are derivative in this sense. Our propositional attitudes and our perceptions, for example, do not seem to derive their representational contents from other states by the use we make of them: they have their content originally, not in virtue of being interpreted by some other mental state.

Many things may thus be said to have content, but most of them do so indirectly: they have content in virtue, for example, of having been produced in a certain way or with certain intentions, or of standing in some relation to other things that have content. The most important such relation is that of some things expressing other things. It is in virtue of expressing my beliefs that my utterances have content, and – subject to certain constraints – the beliefs expressed determine what content they have.

That some representational properties are exemplified intrinsically by some things follows from the following argument:

(i) Some things have representational properties.
(ii) If something exemplifies a representational property extrinsically, it does so in virtue of there being something else that bestows it with this representational property.

---

13. It is only as an analysis of intentionality, not of representationality, I think, that Aristotle's theory of thoughts being likenesses of objects has any plausibility .

14. This has been particularly stressed by Charles W. Morris: "The properties of being a sign, a designatum, an interpreter, or an interpretant are relational properties which things take on by participating in the functional process of semiosis." (1939: 82)

(iii) In order for something to bestow something else with a representational property, the first thing needs to exemplify this representational property itself.

(iv) The transmission of representational powers can neither go on forever, nor go in circle: it must be started by something.

(v) A thing that has a representational property that is not bestowed upon it by something else exemplifies it intrinsically.

Take the representational property *that rabbits are present*, exemplified by my utterance of "Lo, a rabbit". It is exemplified extrinsically: my utterance could be just as it is by itself and express another belief, or no belief at all. By (ii), this representational property is bestowed upon my utterance by something else – something which, by (iii), has it itself. By (iv), we conclude that the regress must stop, and by (v) we know that it must stop with something that *intrinsically* means that rabbits are present.

Representational properties like *meaning that Fa*, *representing a to be F* or *thinking of a as F* are intrinsically exemplified by some thing x iff x exemplifies the property independently of how matters stand with respect to other things than x – no further properties have to be exemplified for other things for my thought, e.g., to represent a to be F.

Even when they are exemplified intrinsically, however, representational properties are still relational: they connect their bearers to the things they are about. If my thought, for example, represents a to be F, it stands in the relation of aboutness to a and in the predication relation to the universal F. It is in virtue of these relations that my thought can stand in for a's being F, and be in some sense a substitute of this external fact.[15]

1.2.4 Correctness and Truth

By being tied to correctness and truth respectively, intentionality and representation differ quite markably in what may be called their 'metaphysical status'.

1.2.5 Indeterminacy of Content

We have, I hope, motivated the theoretical need to proceed (i), via intentionality, from accuracy conditions of pictorial or sense-datum representations to correctness conditions of attitudinal representations and (ii), via representation, from such correctness conditions of property ascription or 'feature-placing' to the truth-conditions of full-blown predications. In this section and the following, we have to pay attention to the degree of indeterminacy of 'content' that is thereby introduced.

Distinguishing correctness from truth allows distinguishing two sources in whcih indeterminacy in (the attribution of) content to representational states may arise. There may be different, equally good (i.e. bad) candidates to account for their intentionality, different things, features or facts they may be said to be about. Many mental states, while they are directed and directed towards things (features, fact) of a certain sort, may not meaningfully be said to be directed towards anything in particular. In addition, and independently of this indeterminacy, there may be no meaningful way of attributing to them one particular, rather than another quite different, set of truth-conditions. While they do impose a condition on the world for their truth, the condition does not correspond to any determinate way or set of ways the world might be.

The Determination of Aboutness

Even though intentionality (directedness, aboutness), as has been argued, is a genuine feature of many of our attitudinal reactions to the world, it is in many cases very much indeterminate what a mental state is directed towards.

FOR EMOTIONS. PROBLEM OF SUPERVENIENCE For emotions, the problem of indeterminacy arises in the following way. When I fear the lion, my fear is directed towards and a reaction to (present or absent) danger, but is appropriate only if the lion really is dangerous. Lions, however, are not dangerous *tout court* – they are

---

15. Different accounts of this relation of standing-in have been proposed, from Aristotle's 'being-a-token-of' – "It is not possible to converse by bringing in the objects themselves, but instead of the objects we use words as tokens", *Sophistici Elenchi* 1, 165a6-8 – to the scholastic modes of objective existence.

dangerous in virtue of exhibiting other properties, such as being hungry and aggressive, having sharp teeth, not being separated from their putative prey by e.g. a cage etc.

One problem with non-existent things is that there is a huge (and moreover: an indeterminate) number of them. We all have heard of Pegasus, the winged horse that was, according to Greek mythology, captured by Bellerophon. How many winged horses do not exist? What about the Easter Bunny?

DIRECTED, BUT NOT TOWARDS ANYTHING IN PARTICULAR

I OWE YOU A HORSE

The Attribution of Representational Content

The indeterminacy introduced in the second step is of a different kind, both with respect to its nature and to its theoretical explanation.

The problem of the indeterminacy of representational content arises from the combination of two factors. On the one hand, the world makes more distinctions than we do; this follows from the fact that at least in a large number of cases, we do not 'make' distinctions at all, but rather latch onto distinctions that are already there, privileging some by marking them representationally, plastering over many others. On the other hand, what distinctions are relevant varies with time, across people, even with contexts: even with respect to our own representational repertoire, we almost always underexploit its capacity to make distinctions.

When we attribute representational content to some state and cannot specify it in terms of representational content possessed by some other state, the conditions we impose on the world as those under which the state is true are, by the second factor, coarser-grained as other conditions we could also impose. By the first factor, these other, more fine-grained conditions do generally pick out different ways in which the world could make the state true. With what justification, therefore, do we take the broader and not the more narrow conditions to specify the content in question?

Within a theory of mental representation this problem arises with respect to the question how exactly mental representations acquire their truth conditions. One superficially tempting proposal is that the representational content of an internal representation is determined by simple counterfactual co-variation relations, perhaps backed by causal laws and past causal dependence (Stampe 1977; Dretske 1981). Causally-backed co-variation relations are ubiquitous: tree rings causally co-vary with the age of the tree, smoke causally co-varies with fire, and utterances of 'Fire!' causally co-vary with the presence of a speaker at the time and location of the utterance. It is therefore widely agreed that simple causal co-variation relations cannot be what determines the representational content of mental states. First, causal co-variation cannot adequately account for the possibility of error. A belief or perception can misrepresent the world – but a tree ring cannot misrepresent the age of the tree. In general, if we identify that which is represented by a given state type with whatever causally co-varies with it, then the existence of the mental state will guarantee the existence of the thing represented. A second related problem is that such 'natural' representation cannot explain the determinacy of representational content. Your 'horse' concept is accurate of all and only horses, but its occurrence in your thinking causally co-varies with many other things, including the proximal retinal stimuli that lead you to identify horses, as well as many non-horses in the distal environment. So the worry is that causal co-variation accounts of mental representation will take 'horse' to represent not the natural kind HORSE, but the disjunctive kind HORSE-OR-HORSEY-IMAGES-ON-THE-RETINA-OR THOUGHTS-OF-COWBOYS-OR-COWS-ON-A-DARK-NIGHT-OR-DONKEYS-IN-THE-DISTANCE-OR-... (Fodor 1984). If we want to explain the determinate content and fallible truth conditions we attribute to mental states, we need a less permissive representation relation than mere causal co-variation.

To further illustrate this obvious point, consider a much-discussed example: magnotactic bacteria. Certain anaerobic bacteria possess internal structures – magnetosomes – that are sensitive to magnetic fields. In the bacterium's home environment, these structures normally align with the earth's magnetic field, causing the bacterium to move downwards towards oxygen-poor water, which is essential to its survival. So there is a reliable co-variation – backed by causal laws and selected for by evolutionary processes – between the alignment of the magnetosomes with the direction of a magnetic field, the North Pole, the ocean depths, oxygen-free water, survival-friendly conditions,

and so on. Should we say that the alignment of the magnetosomes in these bacteria is a representation with the content, say, *magnetic north is in that direction*, or *oxygen-poor water is over there*, or *it's safe down there*? The reason not to attribute them such contents is not in the first place their lack of the supposedly necessary conceptual capacities. The main reason is that attributing them such representational properties is not doing any real explanatory work. As many theorists have pointed out, we can give a full explanation of the movements of the bacterium in one direction rather than another by providing a mechanical account of (i) why magnetosomes are sensitive to magnetic fields, (ii) how the presence of a magnetic field controls their alignment within the organism, (iii) how alignment of the magnetosomes affects the orientation of the organism as a whole, and (iv) how the organism's flagella propel it in the direction in which it's oriented. There is no need to mention representational states with intrinsic truth conditions in this explanation – the mechanical account is a complete proximal explanation the organism's reaction to the stimulus.

Not all cases of measurement thus involve representation. Which ones do? With respect to the accuracy conditions invoked by vision science, Burge has argued that they are plausibly construed as being intrinsic to the perceptual states themselves: perceptual states are individuated in part by their function of representing particular distal features of the environment. Moreover, Burge argues that individuating perceptual states in this way is crucial to the explanations of perceptual psychology – vision science, in Burge's view, is organized around explaining how representations of the world are constructed from impoverished sensory input:

> The fundamental mode of explanation in the perceptual psychology of vision is to explain ways in which veridical representations of the environment are formed from and distinguished from registration, or encoding, of proximal stimulation. Veridicality, fulfillment of representational function, is the central explanandum of visual psychology. Illusions are explained as lapses from normal representational operation, or as the product of special environmental conditions. (Burge 2010: 310-311)

Assuming that perceptual states do have intrinsic accuracy conditions, what exactly do they represent? A frog snaps at a passing fly. Assuming that its visual system is representing some feature of the world, what exactly is the content of that representation? Does the frog's visual system represent something as specific as the content that there is a house fly (musca domestica) at location *x*? Or does it represent that there is a packet of frog food at location *x*? Or does it just represent that there is a small, dark, moving object at location *x*? *Pace* Burge, the explanatory function of content attribution itself does not decide among these alternatives. In many cases, the possible psychological explanations themselves compete. Consider alternative explanations of a five-year-old's choice of ice cream over broccoli:

1. The child chooses ice cream over broccoli because in the past she perceived ice cream as sweet and she likes sweet things.
2. The child chooses ice cream over broccoli because in the past she perceived ice cream as having concentrated nutrients and she likes things with concentrated nutrients.
3. The child chooses ice cream over broccoli because in the past she perceived ice cream as containing large proportion of mono- or disaccharides and she likes things containing large proportion of mono- or disaccharides.

While these explanations differ in what they attribute to the child, and what they predict about its future behaviour, there may simply fail to be, not only in practice, but even in principle, any way of singling out the correct one. What is represented intrinsically by whatever ground the child has for her choice may be radically, metaphysically and not just epistemically indeterminate.

1.3 Belief

1.3.1 Representation-as

We have noted above the close kinship of my fear of the lion, whereby I represent it as dangerous, and of my perception of the lion *as* dangerous. Though it should not be identified with my fear (for it may occur without it, as e.g. when I am courageous), such aspect-perception often accompanies and derives from the emotional state.

Such aspect-perception involves a curious doubling, or re-duplication, of content. There is, one the one hand, the (supposedly) real danger in front of me facing me in the shape of a ferocious lion; this is the danger my fear is directed towards and intentionally represents. On the other hand, danger is also an aspect of how the world is given to me: (the feeling of) danger is an aspect of my situation, a feature of how *I* am when faced with the lion. Though not much more than a metaphor, representation-as brings out this double aspect: I represent the dangerous lion as dangerous, and thus correctly.

In Aquinas' theory, perception, the central and paradigmatic case of representation, is the combination of two processes, which together constitute the uptake of form without matter. There is, (i), the **the getting of the** (inner) **form**: by the application of a representational faculty to something out there in the world, something inner is created (the form of the thing outside in its 'intentional existence', ie. a structural property of the brain); by being in such a state, we locate ourselves with respect to things outside of us. There is also, (ii), the **the seeing of the** (outer) **form**: our brain being in such a state causes us to see the form, ie. to reach out in more or less successful ways to the things we take to be of this form; as a result of this activity, the form out there gets to be seen. It gets to be seen as such-and-such, and in certain ways, in virtue of the way the seeing is done. In different types of representation, the importance of these components varies. In sensation, the first of these components is much more important than the second; the act-aspect is mere awareness, which, as Brentano correctly recognised, is often already contained in the representational state. With imagination, emotion, judging, thinking, and mere supposing – states which do not only have representational properties, but exhibit intentionality as well – the second, mind-to-world aspect becomes more important. Even with such 'higher' forms of cognition, however, the first component never completely disappears: all such mental acts are properly understood as reactions.

This basic structure of representation-as is perhaps most clearly discernible in the case of emotion: when I am afraid of the lion, I locate myself with respect to it (putting it, so to say, on my mental map of the world), but I also reach out to an axiological aspect of the world, the lion's being dangerous. It is in virtue of this formal (or, in Brentanian terms, 'secondary') object of my fear that I represent the dog as dangerous (Mulligan 2006a, b, 2007; Teroni 2007). This representation, however, is not just grounded in the 'mode' of my entertaining a certain content, or the 'attitude' I have towards the direct object of my state, the lion.[16] It is, rather, a reaction to an independently given feeling of value, which reveals to me the dangerousness of the lion. Only by construing the emotion as a reaction, can we explain how it contingently has the valence it has, i.e. why, for non-perverted people, fear is not a pleasant emotion to have, and why being afraid feels a certain way.

That the attitude I take towards the cognitive base comes with a feeling explains, I think, the phenomenology of the mental. This 'inner side' of our mental states, which enables and sometimes constitutes our awareness of them, is a matter of how they feel. To account for how they feel may be one way of cashing out what it means to say that 'there is something it is like' to be in them, though the latter notion is so slippery that it is hard to be sure. That my taking up an attitude towards a representational state that relates it to a formal object is quite unmysterious and ubiquitous phenomenon; while it requires, as its ground, that the attitude has a determinate nature, it does not require either phenomenal content nor autonomous phenomenal properties.

CONSTRUCTION OF THE PERCEPTUAL OBJECT (ie the thing I see as being so-and-so) – IS LATER, IN INTENTIONAL OBJECTS

---

16. As Teroni and Deonna would have it, cf. their 2008; 2012; 2015).

Belief in appearances

Hedging Russell's Retreat

When things appear to us a certain way, we may report this by saying that they have, or produce in us, such-and-such appearances, whether or not they really are as they appear. It is often thought that this manoeuvre (the 'Russellian retreat') increases justified confidence: Although we (often) do not know how things are, or are 'in themselves', we may at least be confident that they appear to be a certain way. The epistemological gains, however, come at a metaphysical price: what are these appearances, and what in their nature explains that they are more accessible than what they are appearances of? In my contribution to the conference, I would like to explore an alternative picture according to which belief how things appear is in some sense weaker, not stronger, than belief how things are.

Let us start with an assertion, by someone at a certain type, pointing to a sheet of paper in front of her, of

(1)   This is red.

This assertion expresses a belief with the content that the sheet in question is red and is true iff the sheet is red. Suppose now that the same person, in the same circumstances and with respect to the same sheet, says

(2)   This looks red.

I will suppose, for the moment, that the person could have made an assertion relevantly equivalent to (2) with any of the following: "This looks red to me.", "This seems / appears red [to me].", "This has a red look [for me].", "Of this object, one has [I have] a red impression. I will also assume that (2) is true iff the sheet looks red, i.e. iff the sheet has a property, *looking red* (or perhaps: *looking red to someone in such-and-such circumstances*), we may call an "appearance property".

But what is *expressed* by (2)? It is commonly assumed that (2), like (1), expresses a belief, differing from (1) only with respect to its content: while (1) ascribes to the sheet the property *being red*, (2) ascribes to it the different property *looking red*. On most views of such properties, application conditions for their predicates are *easier* to fulfill than the application conditions for the predicates they embed: to be justified in (1), I have to be justified in thinking that the sheet is *really* red, while justification fo (2) is easier to come by.

I want to epitomise this view by the label "Russell's retreat". Russell's retreat is the view that in asserting not just (2) but (1) as well, I take a certain type of epistemic risk (perhaps because (2) expresses some defeasible evidence for (1)): when challenged with respect to (1), it may make sense to 'fall back to', or 'retreat' to the *weaker* assertion (2).

In contrast to this picture suggested by Russell's retreat, I want to suggest that (2) is expressive not of belief, but of another state of mind – of conjecturing, taking it to be possible, or even of surmising. As my starting-point and motivation, consider what Sellars says about "looks"-statements in *Empiricism and the Philosophy of Mind*:

> …to say that a certain experience is a *seeing that* something is the case, is to do more than describe the experience. It is to characterize it as, so to speak, making an assertion of claim, and – which is the point I wish to stress – to *endorse* that claim. […] …the statement "*X* looks green to Jones" differs from "Jones sees that *x* is green" in that whereas the latter both ascribes a propositional claim to Jones' experience *and endorses it*, the former ascribes the claim but does not endorse it. (Sellars 1997: 39, 40-41)

If (2) does not endorse the claim that the sheet is red, what does it do? Sellars says that it "report[s] the fact that my experience is, so to speak, intrinsically, *as an experience*, indistinguishable from a veridical one of seeing [that the sheet is red]" (1997: 39, 40-41). This may well be. But what is *endorsed* by (2)? Certainly not what is reported: asserting (2) I am not endorsing any claim about what would be indistinguishable to what, under such-and-such circumstances and for such-and-such subjects.

What I am endorsing in asserting (**2**), plainly, is *that this looks red to me*. Does this bring us back to the appearance-properties with respect to which we are supposed to enjoy some epistemological privilege? Not necessarily, I want to argue. There are three question I would like up in turn:

1. Does the *content* of (**2**) (or rather: of the mental state expressed by (**2**)) involve an appearance-property?
2. Do the *correctness conditions* of (**2**) involve an appearance property?
3. Do the *truth conditions* of (**2**) involve an appearance property?

Construing "seems" as a (type of) copula, it seems to me, gives us the material for a negative answer to the first question. Motivation for a negative answer to the second question may be found, I want to suggest more tentatively, may be found in Kant's distinction between judgments of perception and judgments of experience. Even more speculatively, grounds for a negative answer to the third question may be found (or at least looked for) in a certain view about the mind of the Pyrrhonian skeptic or, alternatively, in a view about what God has to know to know everything.

"Seems" as a copula

Broad (1923: 237) distinguishes between what he calls the "Multiple Relation Theory of Appearing" and the "Sensum Theory". According to the former, which he finds suggested (though not worked out) in the writings of Dawes Hicks and G.E. Moore, "*appearing to be so and so* is a unique kind of relation between an object, a mind, and a characteristic". "Seems" is treated as a "mode of copuation", "allow[ing] Moore to acknowledge the possibility than an object of awareness might be united with [red] only relative to my current experience" (Martin 2015: 150). According to this 'theory', the way in which a region of space has a colour (is "pervaded" by it) is a three-place relation, "involving the pervading colour, the pervaded region, and another refion which we might call the "region of projection"." (1925: 162).

Judgments of perception and judgments of experience

cf hand1-existence-KantianCopula.tex

Bader: it's a distinction along the logical vs real use, ie between general and transcendental logic, in terms of how the faculty is employed, form and matter of representation

my view: instead of assimilating hallucinations to illusions (seeing a region of space to be occupied by pinkness and elephanthood), assimilate illusions to hallucinations (you think there is a FG object, you posit it in a thetic judgement, but relatively so (as is done in problematic judgments of possibility).

Sweet and makes-an-impression-of-heaviness are both response-dependent

In some works, though not in the second edition of the *Critique of Pure Reason*, Kant makes a distinction between two types of judgment, i.e. two ways in which a predicate and a subject concept can be mentally combined in an act of endorsement.

In contrast to judgments of experience (JE), judgments of perception do not require a pure concept of the understanding, but only the logical connection of perceptions in a thinking subject (*Prolegomena* §18). The step to the JE is that "beyond the perception is added the understanding's concept of a cause, which connects *necessarily* the concept of sunshine with that of heat, and the synthetic judgment becomes necessarily universally valid, hence objective, and changes from a perception into experience" (§20, fn.). In the Mrongovius Metaphysik (early 1780s), Kant says that for a judgment to have more than subjective validity, the sequence of perceptions [expressed by it, PB] "must be determined according to rules, i.e., be necessary", and this in turn means it must be necessitated by being grounded in how the world objectively is (AA 29: 815-16).

Not all pairs of concepts can be combined in both types of judgments: "body" and "heavy" can be combined both in a JE, "the body is heavy", and in a JP, "when I carry the body, I feel it to be heavy [einen Druck der Schwere]", but "sugar" and "sweet" only get together in a JP, "the sugar is sweet". It is thus clear that Kant did not have a syntactic difference in mind. Though the details are murky, the crucial difference seems to lie in the

claim to validity: judgments claiming objective validity are JE, judgments claiming only 'subjective' validity are JP. But what could subjective validity be?

Are judgments of perception merely problematic?

Are judgments of perception synthetic existence judgements?

The question of parallelism:

- "…the same function which gives unity to the various ideas in a judgement also gives unity to the mere synthesis of various ideas in an intuition" (A79-80/B105-106)
- "Imagination is intuition even without presence of the object, and the object is then called a phantasm, which can be a production (invention) or reproduction (recollection) of an intuition that was had previously." (R6315, 18:618)
- Can we predicate of intuitions? "Da keine Vorstellung unmittelbar auf den Gegenstand geht, als blos die Anschauung, so wird ein Begriff niemals auf einen Gegenstand unmittelbar, sondern auf irgend eine andre Vorstellung von demselben (sie sei Anschauung oder selbst schon Begriff) bezogen. Das Urtheil ist also die mittelbare Erkenntniß eines Gegenstandes, mithin die Vorstellung einer Vorstellung desselben. In jedem Urtheil ist ein Begriff, der für viele gilt und unter diesem Vielen auch eine gegebene Vorstellung begreift, welche letztere dann auf den Gegenstand unmittelbar bezogen wird." (A59/B93)
- "We know any object only through predicates that we can say or think of it. Prior to that, whatever representations are found in us are to be counted only as materials for cognition but not as cognition. Hence an object is only a something in general that we think through certain predicates that constitute its concept. In every judgment, accordingly, there are two predicates that we compare with one another, of which one, which comprises the given cognition of the object, is the logical subject, and the other, which is to be compared with the first, is called the logical predicate." (R4634, dated to 1772 to 1773)

Living by appearances

Pyrrhonian sceptics faced a similar question. Diogenes Laertius IX 103-4 objects to skeptics professing to live without beliefs that they have to "reject what is apparent". Sextus replies:

> …we do not overturn anything which leads us, without our willing it, to assent in accordance with a passive appearance – and these things are precisely what is apparant. When we investigate whether existing things are such as they appear, we grant that they appear, and what we investigate is not what is apparent but what is said about what is apparent – and this is different from investigating what is apparent itself. (*PH* I x, 2000: 8)

Beyond the act/object distinction: Look-statements as statements about looks

this is first done in Read, and then in Brentano then Russell (give quotes) then Moore: "…whatever be its nature, the entity which is experienced must in all cases be distinguished from the fact or event which consists in its being experienced" (Moore 1914: 169 of reprint) general act/object distinction for any act is already in Moore, Refutation of Idealism

Martin, What's in a Look: Even though he is right that there is a non-evidential, non-epistemic and non-comparative use of "looks", Jackson is wrong about what he takes to be paradigm cases of the 'phenomenal' use. Used phenomenally, "looks" is characterising not the thing that looks a certain way but instead the look the thing has. The clearest such cases concern other sense modalities: "This wine tastes sweet" and "He smells rank" do not (or at least not primarily) qualify the wine or the gentleman, but rather the "manner of appearance". We find the same structural ambiguity in "The spread looks splendid", which may mean (i) that the look of the spread is splendid, or (ii) that the spread has a look that is relevantly similar to the look of splendid things. Jackson mistakenly reads "This looks red" as forcing upon us reading (i), attributing redness to a 'pure visible', i.e. a sense-datum. But reading (ii) is much more natural, and less revisionary.

[Johnston](2004: 131) (file, in file ([Byrne & Logue](2009)) gives an act/object account of hallucination: hallucinations give us de re knowledge, and not of qualities of experience types BUT (PB) this presupposes that the only way an experience can give us knowledge about a quality is either by having it or by representing something as having it (in this latter case it wouldn't be de re). ENCAPSULATING may be another way.

CF ALSO PRICHARD'S CRITICISM OF THE REIFICATION, AND ITS RELEVANCE TO KANTIAN ASPECTS (AND THE KANTIAN FALLACY)

Cf interesting letter by John Cook-Wilson to G.F. Stout in 1904:

> If we perceive some property of an object, there is presupposed on the one hand the property of the object as existing on its own account whether we perceive it or not; and as distinct from this, our act of perceiving or recognizing the nature of this property.
>   This latter, the subjective act of ours, is sometimes spoken of from the side of the objects as the *appearance* of the object to us. This 'appearance' then gets distinguished from the object, and that in itself is justified in so far as our subjective act of recognition of the object's nature is not the same as that nature. But next the *appearance*, though properly the appear*ing* of the object, gets to be looked on as itself an object and the immediate object of our consciousness, and being already, as we have seen, distinguished from the object and related to our subjectivity, becomes, so to say, a merely subjective 'object' – 'appearance' in that sense. And so, as *appearance* of the object, it has now to be represented not as the object but as some phenomenon caused in our consciousness by the object. Thus for the true appearance (= appearing) to us of the *object* is substituted, through the 'objectification' of the appearing as appearance, the appearing to us of an *appearance*, the appearing of a phenomenon caused in us by the object. (The thing to emphasize on the contrary is that the so-called appearance is the appearing of the *object*, that is, we have the nature of the object before us and not only some affection of our consciousness produced by it.) ([Cook Wilson](1926): 796–797)

The so-called appearance is the appearing, not the object!

Adequatio intellectus et rei

In an intringuing article, John [Hawthorne](2005) explores ways in which an omniscient being could know what we express using vague predicates. He considers the possibility that God, facing a Sorites series, becomes quite difficult for us to interpret:

> I say to God: "Ok, I'm going to take away one hair at a time and at each step tell me if the person in front of me is not bald." God replies: "Ok I'll draw a picture of a bald person if and only if the person is bald." As I remove progressively more hairs, God draws increasingly hair-free depictions. After a while, it will be plausible to say that God has stopped drawing pictures of people who are definitely not bald. And a while after that, it will be plausible to say that God has started drawing pictures of people who are definitely bald. As for drawings made during the intervening period, we will say on each occasion that God has drawn a picture of someone who is neither definitely bald nor definitely not bald. The utterance type which is used to say that the person is bald is the act of drawing a bald person. Obviously there will be cases in which it is indefinite whether the utterance is performed.

1.3.2 Different Types of Normativity

We have seen, in sct. 1.1.4 that the intentionality of perception may be explained in terms of its accuracy conditions,

We have further seen, in sct. 1.2.1, that we may explain the representational character of emotions in terms of formal objects. They are 'directed at' both a material object (the dangerous dog in front of me) and a formal object (the dog being dangerous to me in this situation) in virtue of

We have seen above that our epistemic practices are governed by a special kind of norms and evaluable with respect to a special kind of values – 'epistemic' norms and values respectively. It is now time to complicate this picture.

As avid lovers (and producers) of distinctions, philosophers tend to take a quite callous, 'the-more-the-better' attitude towards them: rare a philosopher who cannot be argued into making yet another distinction. Sometimes, this is wishful thinking: distinctions are not just between things, but also *within* some envelopping category, and thus making them brings with it the obligation to explain what the commonality between the things distinguished consists in. Unfortunately, the menu of options is quite limited: ambiguity, polysemy, specification on the linguistic side, determinate/determinable, species/genus or perhaps kind/sub-kind as metaphysical options. Even when none of the available options looks plausible, distinctions are still being made.

Distinguishing different 'types' of normativity well illustrates this conundrum. When we say, of Sam, that he is epistemically but not morally good, for example, are we making a distinction such as between, e.g., him being *good-at* reasoning but not so *good-at* acting (or perhaps: behaving morally?), or rather that he is a better reasoner than a moral agent, or rather that he is "good" in the epistemic sense but not in the moral sense? The most popular way of drawing the distinction introduces further complications: saying that his reasoning, but not his acting deserves praise, not only forgets that his reasoning is part of his acting, but also assumes that it is the same thing ('praise') he deserves for the one but not the other. Not only do we rarely praise people for their reasoning, but even when we do, we do not generally exhort them, but simply acknowledge that their reasoning is well, i.e. competently (and sometimes even skillfully) done.[17] Whether they *should* do it at all, or even should do it well if they do it at all, is left open by our registration of its quality.

It is in this sense that we should understand our talk of "appropriateness", "correction" and "truth": an appropriate reaction is not a reaction we should have, or are obliged to have, but a reaction that is understandable and which, for this reason, we are not criticisable to have. "Appropriate" epistemic actions, however, differ from other non-criticisable actions (such as hobbies) in another way: they are reactions, i.e. responses to outside triggers. For such a reaction, we may be criticisable in two ways: for having a reaction at all, and for having the type of reaction we have. We thus get two different types of epistemic norms and the appropriateness of the reaction is thus the joint outcome of two factors:

- it must be caused by (and justifiable in terms of) an appropriate *cognitive base*, which is directed at the representational / direct objects and represents them as being a certain way;
- it must be accompanied by an *affective feeling* or *cognitive attitude* that is the right kind of reaction to the right kind of intentional / formal object.

Because we account for the ephemeral or non-canonical objects in terms of seeing-as, we may characterise them as manifestations, in virtue of the perception of which we perceive the things they are manifestations of, avoiding postulating necessary connections between distinct existences. My relation to the value I see exemplified is passive, reactive, mind-to-world rather than world-to-mind. Our access to value thus is Schelerian Wertfühlen rather than neo-sentimentalist 'construction'.

Representational properties, being intrinsic, are not had in virtue of things outside the entities that have them. They are relational, however, in two ways: in virtue of reaching out to their direct objects, if any, and in virtue of being representations of these objects *as* such and such. It is this second feature, representation-as, which accounts for perspectivity, and the perspectivity is located in the mode of the activity described under (ii) above, in the way it is done. The content-conditions for representation-as have a two-fold structure: To represent something *a* as *F*, I must be *appropriately related* to a item external to the representation (this may be a 'proposition', a 'state of affairs', a perceptual situation, an object, a plan) and my standing in this relation to the external item must be an appropriate *reaction* to it (serious, non-lucky, veridical, non-perverse, not practically irrational). We may call the first item the **cognitive base**, and the second the **attitude** taken up towards it.

The appropriateness of the reaction is thus the joint outcome of two factors:

- it must be caused by (and justifiable in terms of) an appropriate *cognitive base*, which is directed at the representational / direct objects and represents them as being a certain way;
- it must be accompanied by an *affective feeling* that is the right kind of reaction to the right kind of intentional / formal object.

---

17. I am not sure that such acknowledgment deserves to be called "praise", but intuitions (and cultural practices) may well vary here.

The two factors are intertwined in that the intentional object is the representational object as perceived a certain way.

Under what circumstances is the reaction to the cognitive base appropriate? The answer exhibits again the double structure sketched above: the reaction must be caused by (and justifiable in terms of) its base, and it attitude must match its formal object, i.e. the axiological property it attributes and in terms of which it is individuated. The double structure of representational states, reaching out to something beyond themselves and at the same time situating us with respect to it, allows for two types of failure to meet the characteristic content-conditions for the respective state: either the failure concerns a mislocation of ourselves (wrong direct object), or a constitutive condition on the attitude in question fails to be met (wrong formal object). What this means concretely depends on the form the content-conditions take:

1. Judgements have success-conditions: My judgement of a to be F is successful iff I succeed in referring to a and in predicating F of it. Two types of failure are possible:
   (a) Wrong direct object: I do not refer to a (but to something else or to nothing at all) or do not predicate of it to be F (but to be something else, or nothing at all).
   (b) Wrong formal object: I do not succeed in commiting myself to a's being F, but perform some other action or no action at all.
2. Beliefs have truth-conditions: My belief that a is F is true iff Fa. Beliefs are criticisable on two grounds:
   (a) Wrong direct object: it is not true that a is F.
   (b) Wrong formal object: my belief falls short of knowledge, i.e. has not the right genesis to be a manifestation of my desire for truth.
3. Perceptual experiences have veridicality-conditions: My perceptual experience of a as F is veridical iff I am visually presented with a and see it as F. Two types of perceptual 'error' may be distinguished:
   (a) Wrong direct object (illusion): While I do have a perceptual experiences as of a being F, I am either not visually presented to a or do not see it as F.
   (b) Wrong formal object (hallucination): My state cannot be characterised as one of perceiving.
4. Emotions have appropriateness-conditions: My emotion of type T towards a is appropriate iff a exhibits the formal object of emotions of type T. Two types of error:
   (a) Wrong direct object: My fear is not directed at the dog or the dog, at which my fear is directed, is not in fact dangerous.
   (b) Wrong formal object: I see the dangerous dog, but I don't see it as dangerous. While I am struck by the right value (dangerousness), I don't react in the right way (by fear).
5. Desires have satisfaction-conditions: My desire to ϕ is satisfied iff I want to ϕ and I ϕ. Again, two types of error are possible:
   (a) Wrong direct object (akrasia): I can ϕ, but do not ϕ.
   (b) Wrong formal object: I ϕ, but I don't do it as a my plan. While I realise my plan, I don't react in the right way to my ϕ-ing, i.e. I do not see it as the satisfaction of my desire.

Whenever there is a mismatch between attitude and the intended formal object, I dishonour a certain value. Because the attitude itself is voluntary, and can be withhold, the standard of correctness is internal to it and conditional on my having the attitude. This explains why the normativity of the mental, contrary to what is often held, is never a matter of some mental state being obligatory (under some circumstances), but exclusively a question of its being permissible (or perhaps good) to be in such a state (cf. Blum 2011). That emotions and beliefs, e.g., are optional in this sense means that they are only internally criticisable, and that the values that justify the (conditional) norms governing them are in some sense 'internal' to them.

For belief and emotion, the falsity of prescriptivism is demonstrated by the blamelessness of two familiar philosophical characters, the Sceptic and the Stoic respectively: the Sceptic acknowledges things as true, but does not react with belief, 'stays with appearances', even though belief would be appropriate. She is blameless, even though perhaps criticisable for not doing as well as she can (though such a criticism certainly requires substantial philosophical argument). The Stoic, on the other hand, acknowledges things as offensive/dangerous etc. but does not react with anger/fear etc – he stays with 'feelings of values' (cannot avoid being struck by them), but does not

respond with an emotion that would be appropriate. From the point of view of a long and honorable tradition, the Stoic is not only blameless, but indeed praisworthy – certainly not a possibility an account of the nature of emotions should rule out!

While professing non-sense, wishful thinking, hallucination, psychopathic 'emotions' and the 'desire' for water of the hydropic are not judgments, believings, perceptions, emotions and desires respectively, they still stand under (and are criticisable in terms of) the same norms. This is a difficulty for disjunctivists who cannot explain it just by the subjective indiscernibility of the two states. Formal objects fill in the lacuna: even though hallucinations are not 'failed' perceptions, but something different altogether, they are still subject to the same norms because they have the same formal object as perceptions. In the same way you can be afraid of spiders, but only if you are afraid of them as dangerous, you can hallucinate unicorns, but only if you see them as existing.

Lessons for normativity

For belief and emotion, the falsity of prescriptivism is demonstrated by the blamelessness of two familiar philosophical characters, the Sceptic and the Stoic respectively: the Sceptic acknowledges things as true, but does not react with belief, 'stays with appearances', even though belief would be appropriate. She is blameless, even though perhaps criticisable for not doing as well as she can (though such a criticism certainly requires substantial philosophical argument). The Stoic, on the other hand, acknowledges things as offensive/dangerous etc. but does not react with anger/fear etc – he stays with 'feelings of values' (cannot avoid being struck by them), but does not respond with an emotion that would be appropriate. From the point of view of a long and honorable tradition, the Stoic is not only blameless, but indeed praisworthy – certainly not a possibility an account of the nature of emotions should rule out!

1.3.3   Supposing, Holding True and Judging

To distinguish, as we have done, intentionality and representation is not to deny that there are any explanatory links between the two. While the modern orthodoxy generally explains intentionality by representation, cooking up an 'intentional object' for any way I can represent the world to be, the Aristotelian tradition, rightly understood, proceeds in the other direction, explaining representation by intentionality.

supposing opining conjecturing holding true believe believe truly know judge

It's not natural language analysis!

Second role of intentional objects for belief: In belief, intentionality does not just determine aboutness, but also the directness to truth; so it explains why belief states have truth-conditions AT ALL..

1.3.4   From Belief to Judgement

MAIN POINT: BELIEF IS ALREADY ASSENT-INVOLVING (distinction between believing and entertaining ('opining' perhaps in Sellars and Williamson) is crucial to my criticism of the knowledge rule of assertion diagnosis of Moore's paradox).

Contemporary epistemology, analysing knowledge in terms of belief or rather belief in terms of knowledge, uses a rather anemic notion of 'belief', according to which not much more is required for believing that p than to have it in your mind, but not as a question. Earlier theorists, in sharp contrast, used the rather pregnant notion of "judgement" to pick out the primary bearer of justification.

Like perception and emotion, belief is a reaction: the representational state is an entertaining of a thought, and the mental act is a type of Stoic assent (in the case of belief; dissent in the case of disbelief). To believe, Augustine says, is 'cum assensione cogitare' (de praed. sanct. 2 5) (cf. Mulligan 2007: 218). In the same way that thinking, by the 'affirmative' mode, gives beliefs, it may, by the 'interrogative' mode, give questions, and by the 'mode of conjecture', suppositions. What is belief a reaction to? Brentano may help us here, drawing our attention to the possibility that inner representation is containment (i.e., works roughly in the way in which a quotation name represents what it quotes by surrounding it by quotation marks). Quoting *De Anima* III. 2, p. 425, b, 12, he remarks that a doubling of

intentional objects is only needed if we individuate them by the representational states, but not if we individuate them (correctly) by the mental acts directed at it: "the presentation [Vorstellung] of the sound is connected with the presentation of the presentation of the sound in such a peculiarly intimate way that the former, by obtaining [indem sie besteht], contributes inwardly to the being of the other [innerlich zum Sein der anderen beiträgt]." (bk. 2, ch. 2, §8, 1874: 167) The resulting picture is the following: There is thinking going on, and some of it is partitioned into a belief of a certain content by an attitude of assent the subject directs at its own mental state. Such a belief, however, is not a representational vehicle in the traditional sense, that allows the subject to represent something. Rather, the belief, as a mental particular (and thus distinguished from the state of believing), is itself produced by (and thus posterior to) the representing of form. The attitude of assent directed at the immanent object is at the same time, and by the same act, a representation of itself as a psychological state, i.e. a belief. It thus projects its own form onto an amorphous mental process of thinking. There is a peculiar interweaving ["eigentümliche Verwebung", 1874: 167] of the object of inner presentation (the belief) and the presentation itself (the thinking), and both belong to the same act, the believing (cf. also Burge 1988: 654).

Such a conception of belief sees it as arising out of less demanding, more fundamental contentful states. We progress, as it were, from mere supposing, to opining, to conjecturing, to holding true and only then to belief.

The value of knowledge

Are there epistemic values? The question is ambiguous between (i) are there specific, domain-specific values in the field of epistemology? and (ii) are there values that are also epistemic? The question has real interest only if taken in the second sense. We are not interested in schwepistemic values, for example, where schwepistemology is defined by its aim of maximising the ratio of bullshit among your beliefs. Someone might say that you schwepistemically ought to believe any bullshit whatsoever, but as this does not entail that you ought to believe any bullshit whatsoever, this schwepistemic 'obligation' is of little interest to us. The question then is whether key topics of epistemology, like truth or knowledge, have value.

"The value of knowledge" (or "the value of truth") presents a second ambiguity: in one sense, "value of" is appositive, as in "the value of justice"; in the other, it is a functor mapping knowledge to something else, i.e. its value. Those who think that the value of knowledge can be characterised in non-epistemic terms understand it in the second way. But we should understand it in this way even if we think, as we should, that the value of knowledge is irreducibly epistemic. Knowledge has value, but is not itself a value – it is an epistemic state, not an ideal aim of epistemic states.[18]

What is the value of knowledge, understood along these terms? I will give my answer in three steps:

   (i) why is it good to have beliefs at all, i.e. to have the attitude of wanting-to-know?
   (ii) why is it good to avoid mistakes?
  (iii) why is it good to be able to provide reasons?

For reasons I hope will become clearer below, I think we should strictly separate these three levels. With respect to them, the question in what the value of knowledge consists receives three different answers:

   (i') it's good to want-to-know because this is noble and sets us above the lower animals;
   (ii') it's good to avoid mistakes, because once we have engaged in (i), we should not do it so badly that it would have been better not to engage in the activity at all;
  (iii') it's good to be able to provide reasons because this makes one virtuous.

level 1: why have beliefs at all?

Why is it good to engage in the pursuit of truth? Why shouldn't we not be disinterested in the truth, abstain from forming any beliefs and from making any judgements at all? We certainly could do so, but it would be a bad thing to do. There are instrumental reasons for this: even if a life without belief may realise a certain number of goods,

---

18. Truth, on the other hand, not only has value, but is itself a value. We will come back to this below.

the pursuit of practical goals is certainly facilitated by beliefs. There are, however, also epistemic reasons, which constitute part of the value of knowledge.

It is good to engage in the pursuit of truth because our ability to do so is part of what sets us above the brutes. Even if other animals may be capable of beliefs, the pursuit of truth as such seems unique to humans. It is also their best claim to intrinsic value – we are the rational animals after all. What does it mean to engage in the pursuit of truth? It is to have a sui generis attitude of wanting-to-know. This attitude is sui generis, because it is not the attitude of wanting or desiring to be in some state of knowing, but rather a direct, un-decomposable and primitive relation to some truthbearer.

Couldn't we have beliefs without engaging in the pursuit of truth? Why should it be impossible to have beliefs in the absence of an attitude of wanting-to-know? And even if it is, of what kind is the relevant impossibility? It would be clearly insufficient to point out that a Moore-paradoxical attitude to one's own beliefs ("I believe that p, but I do not think that p is true") is irrational, for this does not yet answer the question why it is good to be rational in that sense. The connection between believing and wanting-to-know must be tighter. Wanting-to-know is constitutive of our having beliefs, which are attempts at knowledge. Knowledge is in that sense explanatorily more basic than belief (Williamson 2000; Hossack 2007).

While you cannot, I would hold, have beliefs without engaging in the pursuit of truth, i.e. without displaying the attitude of wanting-to-know, it is certainly possible to have the attitude but still live a life without beliefs. That was precisely what the ancient sceptics advocated (Burnyeat 1983: 126): they advocated *ataraxia*, a "matter of not worrying about truth and falsity any more" Burnyeat (1983: 121). But because *ataraxia* was supposed to be grounded in a prior recognition of the equal strength (*isostheneia*) of opposed assertions, it grows out of a prior attitude of wanting-to-know. Those who lack the attitude are fools, not skeptics. Wanting-to-know is in this sense prior to belief – it is presupposed by, but does not presuppose it.

The aim of belief, therefore, is not truth, but knowledge. The most basic epistemic value is the value of knowledge, not of truth.

Why is it good to have beliefs? Perhaps surprisingly: because they may be false. That they may be false means that there is no guarantee that our attempts at knowledge succeed. This is the important truth in fallibilism: there is no easy knowledge. Contra Lewis (1996: 419), fallible knowledge is not knowledge despite uneliminated possibilities of error. Contra Cohen (1988: 91), Stanley (2005: 128) and Brueckner (2005: 384), neither is fallible knowledge knowledge of a truth that is not entailed by the evidence. The reason for this is simple: if either of these accounts were true, there would be no fallible knowledge. If I know that p, then surely p is true in all possibilities uneliminated by my knowledge (cf. Hawthorne 2004: 25). If my evidence does not entail what I claim to know, then I cannot taking (what I take to be) my knowledge to be evidence. Even fallible knowledge is evidence, however (cf. Williamson 2000).[19]

level 2: avoiding mistakes

An analogy to house-building may be helpful here: suppose, for the sake of the argument, that it is good to build houses at all. While there is nothing reproachable about living in caves, the cave-men miss something important. This positive evaluation of house-building is implicitly conditional: it is good to build *proper* houses, i.e. houses that to some degree fulfil their essential function of protecting their inhabitants from the elements. The construction of houses that not only fail to fulfil this role, but positively endanger their inhabitants (in the sense that they are less safe within than outside the house), is not just bad, but forbidden.

---

19. Could we at least define fallibilism about justification as the negation of "if S has justification for believing that p in virtue of having evidence e, then e entail p" (Brueckner 2005: 384–385)? I do not think so. Fallibilism about knowledge and about justification is motivated by epistemic humility, ie. the recognition that our epistemic powers are finite and to err is human. Gettier-cases and the problem of induction are not needed for this.

level 3: epistemic mights, not epistemic oughts

To be sure, where evidence is conclusive if available at all, as it is in logic, we have a negative obligation to avoid mistakes: we should not assert that q if we previously asserted that p and that $p \rightarrow q$ and want to stand by these assertions. Even this negative obligation is conditional, however: it applies only to epistemic subject who decided to engage in the pursuit of truth at all.

1.3.5   The Eliminability of Truth

As fearing the lion is a matter of having an attitude to it that is directed towards (and only appropriate for) danger, i.e. a matter of 'taking' the lion as dangerous, so believing that p is a matter of having an attitude to it that is directed towards (and only appropriate for) truth, i.e. a matter of 'taking' that p as true.

Chapter 2

The Myth of the Given

Chapter 3

The Analysis of Knowledge

3.1 What an Analysis of Knowledge Could Be

3.1.1 Philosophical Biconditionals

"Knowledge" has always been one of philosophers' most cherished concepts. Traditionally, the debate has centred about the third clause in the classic definition of "*a* knows that *p*", commonly attributed to Plato's *Theatetus*, though Plato does not there endorse it:

(**JTB**)
    S knows that *p* iff
    (**i**) It is true that *p*.
    ii) *S* believes that *p*.
    (**iii**) *S* is justified in believing that *p*.

3.1.2 Gettier

In 1963, Edmund Gettier published a short note showing that this analysis is inadequate: though the three conditions might be necessary, they are not jointly sufficient (Gettier 1963). A counterexample is the following scenario: I believe that one pupil in my class is French, on the basis of my justified, but untrue, belief that Joan is French. If I have a justification for this latter belief, I have one for the existential generalisation. It may happen, however, that the generalisation is true, while the instance is not. There is another pupil in my class who, unbeknownst to me, is French. It seems false to say in this scenario that I know that one pupil in my class is French, though the three conditions are satisfied.[1]

3.1.3 Dretske and Nozick: Closure and Tracking

Dretske (1970), Nozick (1981), pp. 172–185 (Teil von sct. 3.I)

The common reaction to such counterexamples was to look for a fourth condition, or, equivalently, a third condition which subsumes the old one, while being stronger than it. Various candidates have been proposed. It has, e.g., been required that my belief that *p* should be counterfactually dependent on the truth of *p* (so that, if *p* were not true in relevantly similar circumstances, I would not (or cease to) believe that *p*). Another way of fixing (**JTB**) has been proposed by Alvin I. Goldman (1967). His idea is to require some sort of causal connection between the epistemic state of the believer and the object the alleged knowledge is about.[2]

---

1. It is not essential to the example that it is the falsity of the instance that undermines the generalisation test. Suppose I tell my class that I have been to Mexico (true) and ask them whether I have been in South America (true). They will acquire justified true belief if they are justified in believing that Mexico is in South America.

2. To allow for knowledge about the future, Goldman draws a difference between there being a causal connection between *a* and *b* and *a*'s being the cause of *b* (or vice versa): the former, but not the latter holds if *a* and *b* are effects of a common cause (Goldman 1967: 364).

In 1981, Robert Nozick proposed subjunctive third and fourth conditions, imposing a sort of counterfactual dependency between the truth of the belief and the fact that it is had:

(**NO**)

> *S* knows that *p* iff
> (**i**) It is true that *p*.
> (**ii**) *a* believes that *p*.
> (**iii***) If *p* were not true, *S* would not believe that *p*. (Nozick 1981: 172)
> (**iv***) If *p* were true, *S* would believe that *p* and it would not be the case that *S* believes that ¬*p*. (Nozick 1981: 178)

Nozick calls such counterfactual dependence "tracking" and explicitly links it to there being a *constraint* on our beliefs and the way the world is:

> "A person knows that *p* when he not only does truly believe it, but also would truly believe it and wouldn't falsely believe it. He not only actually has a true belief, he subjunctively has one. It is true that *p* and he believes it; if it weren't true he wouldn't believe it, and if it were true he would believe it. To know that *p* is to be someone who would believe it if it were true, and who wouldn't believe it if it were false. [...] To know is to have a belief that tracks the truth. Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it." (Nozick 1981: 178)

(iii*), however, is too strong: must be relativised to methods of having the belief in question: If Osama had not lived, I could not have used visual recognition to believe that he lived, but I might have believed it anyway; so I will know that Osama lives even if my belief was not essentially based on perception but on pure superstition. On the other hand, it should be possible to know something and it still be true that you'd use a different method if the belief were not true; you know because you see him; the fact that otherwise you would have believed the propaganda does not do away the knowledge.

Dretske didn't consider (iv*); Nozick's argument for its inclusion is that the account should leave room for knowledge of necessary truths. Independently, however, it is also needed to exclude the following case: Known to me, my speedometer is highly reliable whenever I drive at less than 140 kmh and highly unreliable above (as likely as not to indicate speeds below 140 kmh). Travelling at 150kmh, I look at the speedometer, indicating 150 kmh, and come to believe the truth that I am travelling at more than 140 kmh. (iii) is satisfied because the speedometer is reliable below 140 kmh.

As with (iii*), (iv*) has to be relativised to methods of belief formation: you need not only reliability of the method in cases where the belief is false but also in cases where the belief is true:

(**iv****) If *p* were still true under different circumstances and *a* formed his belief by the same method that he actually used, *a* would still believe that *p*.

(iv**), however, is too strong: it excludes the case where you perceive Jesse James and recognise him even though he normally wears a mask – you know it's him even though in many nearby worlds you would not. So we would have to restrict the condition to cases where you actually form a belief on whether *p*. But even this modified condition is too strong: suppose that Jesse James normally wears a Tom Cruise mask and that in many nearby worlds you would falsely form the belief that he is Tom Cruise.

The most popular candidate was to require some kind of *reliable* connection:

> "*A* truly believes that *p*, and his belief [is] appropriately produced in a way such that beliefs produced in that way are generally true, where "appropriately" means that the truth of the belief is not accidental relative to what it is about the way of its production which makes that a generally reliable way." (Williams 1978: 45)

Fred Dretske, in a series of famous articles, has argued that (**K**) is false of all epistemic operators, in particular of "*a* knows that" and "*a* has reason to believe that".[3] After having searched the whole building and not found anyone in it, I may have reason to believe that the church is empty without having reason to believe that it is a church that is empty, though the truth of the latter is entailed and presupposed by the truth of the former and I may know (and a fortiori have reason to believe) that, even if I know that the latter is a consequence of the former (Dretske 1970: 36). Even if I know that zebras are not mules, it does not follow from the fact that I know that the animals I see in the zoo are zebras that I know that they are not mules cleverly disguised to look like zebras (Dretske 1970: 40). He puts forward what has become to be known as a relevant alternatives approach: in order to know that *p*, I do not have to rule out all possibilities in which ¬*p*, but only relevant alternatives in which ¬*p*. In a given instance of (**K**), I may be able to rule out the alternatives relevant to the premiss-es without being able to rule out the alternatives relevant to my knowledge claim stated in the conclusion.

Convincing as these examples are, they have come under heavy attack. G. C. Stine (1976) famously (?? CHECK) argued that the presupposition carried by a knowledge claim to *p* (that ¬*p* is not a relevant alternative) is only pragmatic, not semantic. In his view, it is a presupposition tacitly made by the speaker and not pertaining to the knowledge claim itself (Stine 1976: 255). Pragmatic presuppositions are cancellable[4] and utterances carrying them may be true even if the presuppositions in question are false. So we may either assert the conclusion, despite our committing ourselves thereby to a false pragmatic presupposition, or withdraw our assertion of the first premiss.

The central idea of this line of response to putative counterexamples to (**K**) is to restrict (**K**) to reasoning *within* a fixed context and to take the set of relevant alternatives to be an essential feature of a context (Stine 1976: 256). A basically equivalent move would be to take different sets of relevant alternatives to define different meanings of "know" and its derivatives (Williamson 2002: 9). Relevant alternative counterexamples to (**K**) would then be guilty of a fallacy of equivocation or, at least, an illegitimate change of context.

Nozick's failure to motivate failure of closure is a step back behind Dretske, which argued against closure in (Dretske 1970). According to Dretske, what the sceptic has done is writing large the situation of the zebras:

(3.1)
$$\frac{\text{These animals are not mules, cleverly disguised.}}{\text{These animals are zebras.}}$$

is analoguous to

(3.2)
$$\frac{\text{I am not a brain in a vat.}}{\text{I have hands.}}$$

If the conclusion were false, you could not use the means of achieving knowledge of the premiss you actually used. The grounds I have for the premiss (visual perception say) are not grounds for the conclusion.

Dretske's analysis, contrary to Nozick's, works equally for stronger epistemic operators than knowledge.

3.1.4 Knowledge as a Disposition to Get Things Right

Nozick's notion of tracking and his requirement that there be a law-like dependency of beliefs on the state of affairs they are about will turn out important in the following.

WHAT'S WRONG WITH TRACKING?

---

3. Other epistemic operators include "*a* sees that", "there is evidence to suggest that", "*a* can prove that", "*a* learns (discovers, finds out) that", "in relation to our evidence it is probable that" (Dretske 1970: 32).

4. An example of a pragmatic presupposition is the exclusive reading of some uses of "or". If I tell a child standing in front of an ice-cream shop that it might have chocolate or strawberry flavoured ice-cream, I normally do not intend to offer him the possibility of having both. Adding "or both", however, cancels this presupposition and thus shows that it was only pragmatic, not semantic.

Can we therefore conclude that the subjunctive account is wrong? Not too quick! Note a similar problem in the analysis of dispositions:

x is fragile $\rightarrow$ if x were subjected to moderate forces, x would break
$$Cx\square \rightarrow Mx$$

This is hopeless: right-hand side is neither necessary nor sufficient for a wide class of dispositions (for which we readily imagine interference between the nature of the thing concerned and the bringing about of the conditions necessary for the manifestation of its dispositional properties). Think of finkish dispositions: whether or not the manifestation conditions are sufficient for the manifestation is there contingent on the categorical nature.

We have to distinguish three cases:

1. altering (finkish dispositions)

2. mimicking (under non-normal conditions, the conditional may be true without the thing having the disposition)

3. masking (the glass is packed which prevents the manifestation)

Nozick makes knowledge look more like a disposition than like a state. Dispositions are in those cases close to states where our interest lies with categorical bases, as with colours changing under illumination by white light (finkish dispositions). Generally, however, dispositions are different from states: with many dispositions, our interest lies solely with their manifestations (as with "being courageous"). Call the first kind of dispositions 'Quinean' and the second kind 'Rylean'. Is the distinction between Rylean and Quinean dispositions interest-relative? Can we say that being fragile is Quinean for physicists but Rylean for postmen?

If you are a Quinean, then you use the subjunctive conditional as a reference-fixer; but we would like them to be more than that. Colours can be altered, mimicked and masked, so they are not Rylean; but it might still be the case that there is no interesting physical similarity among green things - are we forced to eliminativism about colours? (two kinds of jade, many kinds of green etc) - what colours are things on Twin Earth, where on the surface everything is the same but the underlying physics is different? Even on Twin Earth, glass still seems fragile. So we'd better not think of the conditionals just as reference-fixers.

Nozick's mistake was to move too quickly to subjunctive conditionals, where he should have stayed with dispositions. To know is to have a true belief formed as a result of an exercice of a disposition of being right (better: an ability to be right) which is not too local. The Jesse James / Tom Cruise case is literally a case of masking: the mask is masking your ability to recognise Jesse James.

Two readings of the Gettier problem:
1. a gettierized belief does not qualify as knowledge because there is a close, epistemically indistinguishable worlds in which the belief is false ('covariationist reading')
2. a gettierized belief does not qualify as knowledge because the features that render the belief true are not suitably related to the features that render it justified ('identificationist reading')

In the second, non-standard reading, the reasons for which the belief is justified mis-identify the truthmaker for it. Epistemic responsibility is met for another reason than the one which is responsible for the truth-conduciveness of the belief.

3.1.5 Epistemic Supererogation

PERHAPS HAVE HERE THE DISCUSSION OF KNOWLEDGE AND LUCK, PRITCHARD

3.2 Knowledge First

While Gettier's 1963 article re-awakened the philosophical zeal to find a reductive definition of knowledge, some authors recently argued that no such reductive definition, in particular no definition of the form of (**JTB**) and (**NO**) may possible be true. Timothy Williamson, in his influential book *Knowledge and Its Limits* (2000), has argued that knowledge is not analysable in terms of belief at all. He argues that no such factorisation of knowledge into non-circular necessary and sufficient conditions is possible, that knowledge has to be acknowledged as a primitive concept. Rather than to define it, philosophers would better concentrate on spelling out its logical behaviour.

What Williamson needs a story about:

1. $Kp \rightarrow KKp$
2. $(K(p) \wedge K(p \rightarrow q)) \not\vdash Kq$
3. Sorites cases with barns, driving slowly from one context to another

3.2.1 Assertion and Evidence

3.2.2 Williamson on the Primeness of Knowledge

3.2.3 Williamson Against Transparency and Luminosity

Main claims of Williamson (2000):

- Knowledge is a factive propositional attitude such that being in a certain mental state s is a necessary and sufficient condition for knowing the proposition p.
- Knowledge is a mental state even though it is not transparent. Transparency: For every mental state s, whenever one is suitably alert and conceptually sophisticated, one is in a position to know whether one is in s.
- Counterargument: There is an asymmetry between knowing whether one believes that p and knowing whether one knows that p: knowing whether one knows that p requires one to evaluate reasons for and against p, whereas knowing whether one believes that p does not. Therefore, knowing that p is not a mental state.
- Reply (Will.): Even if this were so, there is a mental state, namely the one of rationally believing that p, which parallels the case of knowing that p in that knowledge of it requires the evaluation of reasons for and against p.
- Believing truly is not a mental state while knowing and believing are mental states, although knowing necessarily implies believing truly and believing truly necessarily implies believing. The concept [knowledge] cannot be identical with/analyzed by the concept [justified true belief] because the analyzed concept [knowledge] is mental, while the analyzing concept [justified true belief] is non-mental – it contains the non-mental concept [true].

Knowing is the most general factive mental state. Factive mental states are expressed by factive mental state operators (FMSO) with the following syntactic and semantic properties:

1. An FMSO ϕ has the syntactical properties of a transitive verb: it takes a noun phrase in subject position and a that-clause in object position to form a sentence.
2. 'S ϕs that p' is required to have 'p' as a deductive consequence: attitudes are factive iff necessarily one has it only to truths.
3. An FMSO is stative: it is used to denote a state like knowing or remembering, not a process like forgetting.
4. An FMSO ascribes an attitude to a proposition to the subject: 'S ϕs that p' entails 'S grasps the proposition that p'.
5. An FMSO is semantically unanalysable, although they may be syntactically complex as in 'S could feel that the bone was broken', where 'could feel' is syntactically complex but not semantically analysable into the meanings of the syntactical components 'could' and 'feel' - "'could feel' is semantically fused" (37).

Revolutionary as this thesis was in epistemology, it has long ago been endorsed by situation theorists:

> "Common sense says that knowledge is one thing, belief is another. There is no particular reason to think that one is in any sense more fundamental than the other, or that one can be defined in terms of the other. On the other hand, there are obvious relations between knowledge and beliefs. These relations amount to necessary structural constraints..." (Barwise & Perry 1983: 265)

3.2.4   Epistemic Logic

To spell out these structural constraints, we turn to epistemic logic.[5]

The most important assumption underlying epistemic interpretations of modal logic is that knowledge satisfies the following axiom (**K**):

$$(\textbf{K}) \quad \vdash \mathsf{K}p \wedge \mathsf{K}(p \to q) \ \to \ \mathsf{K}q$$

Together with the following closure principle, this ensures that the logic of knowledge is *normal*:

$$(\textbf{Nec}) \ \dfrac{\vdash p}{\vdash \mathsf{K}p} \ \text{Nec}$$

WHAT'S THE CONNECTION BTW THIS AND KP->KKP ?

DISCUSSION OF CLOSURE ABOVE

The problem is that (**K**) is derivable from the application of the following principles to knowledge K:[6]

$$(3.3) \qquad \vdash p \to q \ \Rightarrow \ \vdash \mathsf{K}p \to \mathsf{K}q$$

$$(3.4) \qquad \vdash \mathsf{K}p \wedge \mathsf{K}q \ \to \ \mathsf{K}(p \wedge q)\,[7]$$

The problem then re-emerges even with operators which are clearly 'non-penetrating' (this is Dretske's term for non-distributivity over 'boxed' implication). Take, e.g., "it is strange that". Even if Dretske is right that "a concatenation of factors, no one of which is strange or accidental, may itself be strange or accidental" (Dretske 1970: 32) and the converse of (3.4) is not true, (3.4) seems ok, as does (3.3) (at least for a restricted sense of "it is strange that" not applicable to astonishing theorems). So we would have the strange result that "it is strange that" distributes. This result is counterintuitive, however, only if we do not consider what "$\Box(p \to q)$" would mean in this case. Suppose it is strange that Susan and Bill married each other but it is not strange that Susan married. Do we then have a counterexample to (**K**)? No, because the premiss $\Box(p \to q)$ is false: it is not strange at all that if Susan and Bill married then Susan got married.

JUSTIFY (3.3): on the face of it, it seems stronger than (K): not just closure under KNOWN entailment but under entailment of any kind?

I do not know whether this strategy works in all cases: there is, after all, nothing bizarre to the idea that some strange fact implies some familiar fact and that this implication is itself strange. The point I want to make is just that putative counterexamples to (**K**) have to be chosen carefully so as to render the major premiss as plausible as possible.

There is another reason to be skeptical about counterexamples to (**K**): what the relevant alternatives to consider in the ascription of knowledge claims are depends, as we saw, on context. This context is not just the context

---

5. Do not be afraid of the symbolism in what follows. While epistemic logics are logics in the strict, syntactic sense, we use here their notation (and some notions from their semantics) just for convenience. Cf. fn. and for a translation scheme.

6. We have $\vdash (p \wedge (p \to q)) \to q$, hence, by (3.3), $\vdash \mathsf{K}(p \wedge (p \to q)) \to \mathsf{K}q$. By (3.4), we have $\vdash \mathsf{K}p \wedge \mathsf{K}(p \to q) \to \mathsf{K}(p \wedge (p \to q))$ and chaining the two together we get $\vdash \mathsf{K}p \wedge \mathsf{K}(p \to q) \to \mathsf{K}q$.
Natural language translation:

7. For the knowledge case, (3.4), i.e. epistemic conjunction, has been denied by Devlin (1991: 186). Unfortunately, he does not give any reasons. Provided one is prepared to accept *any* such logical principles governing knowledge, (3.4) seems to me one of the most obvious candidates.

of the ascriber, but includes at least some features of the context of the person to which knowledge is ascribed. It is therefore possible to argue that the skeptical scenario (q) becomes a relevant alternative to consider just by our ascribing to the subject in question knowledge that p and that $p \rightarrow \neg q$. One way to bring this out is to see the non-obtaining of the skeptical scenario ($\neg q$) as a pragmatic rather than a semantic presupposition which may be cancelled and *is* cancelled in instances of (**K**), where the possibility of p is explicitly considered in the second conjunct of the antecedent. Another way is to follow David Lewis in taking the relevant/non-relevant distinction between possible worlds to be subject to pragmatic rules, one of which is the rule of actuality:[8]

> "When we say that a possibility is properly ignored, we mean exactly that; we do not mean that it *could have been* properly ignored. Accordingly, a possibility not ignored at all is *ipso facto* not properly ignored. [...] No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other contexts, if in *this* context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative." (Lewis 1996: 434)

The skeptical scenario, then, is not properly ignored already in the premises of the argument because it is mentioned in the second one.

Note that it would not be an appropriate line of reply to insist, as it is often done, that the subjects of interest to us (the epistemic behaviour of which we are modelling) are logically perfect reasoners, drawing all logical conclusions from what they know. This reply only makes our situation worse, for it justifies (**Nec**) and thus forces us to accept not only distribution over known implication, but even distribution over any logical entailment whatsoever.

Another general strategy may be suggested by the following consideration. (**??**) fails if we impose, as Dretske did, a requirement on knowledge that is close to Nozick's notion of tracking (cf. def. (**NO**)): in order for me to know that p, there has to be a reason R such that my belief is based on the fact that R holds and such that, if p were false, R would not hold. The problem then is that it seems not clear how to interpret, on these terms, the second premiss so as to make the following inference valid:

$$\frac{\begin{array}{l} \text{If } p \text{ were false then } R \text{ would not hold.} \\ p \text{ entails } q. \end{array}}{\text{If } q \text{ were false, then } R \text{ would not hold.}}$$

The situation, however, changes in important respects if we replace the second condition of the tracking account with its contraposition and require that p is true only if R holds. R may then be said to *indicate* that p is true. We then have the following inference, which is valid:

$$\frac{\begin{array}{l} R \text{ would hold only if } p \text{ is true.} \\ p \text{ entails } q. \end{array}}{R \text{ would hold only if } q \text{ were true.}}$$

While this interpretation of knowledge makes (**K**) acceptable, it suggests at the same time that the notion captured is more precisely described as "having the information that" than as "knowing that".[9] It seems plausible to me that the transitivity of "indicates" is just the manifestation of a more widespread phenomenon, called "nesting" by Dretske.[10]

---

8. The others are the rule of actuality (the actual world is never properly ignored), the rule of belief (a possibility that the putative knower believes to obtain is not properly ignored), the rule of resemblance (we never properly ignore just one of two similar possibilities), the rule of reliability (we properly ignore the possibility that reliable mechanism malfunction) and the rule of conservatism (generally ignored possibilities are properly ignored).

9. A somehow intermediate solution, exposed and defended by Duc (2001), as reported by (Gochet & Gribomont 2006: 50–51), would be to change (**K**) for the temporalised $\mathsf{K}(p \rightarrow q) \rightarrow \mathsf{K}p \rightarrow \Diamond\mathsf{K}q$. The resulting logic eschews all seven forms of closure mentioned in the main text and thus seems to rule out too much. It is a further disadvantage of the system, however, that we are not told *what* the agent has to do to eventually arrive at $\mathsf{K}p$. It is not very helpful to be told that "$\Diamond\mathsf{K}p$" reads "sometimes after using rule R, the agent knows p".

10. It is somewhat ironical that Dretske later advocated an even stronger definition of knowledge according to which it distributes – not only over *known* but – over any necessary connection whatsoever.

(**Nec**), indispensable as it is to the project of modelling knowledge with the tools provided by modal logic, imputes *logical omniscience* on the epistemic agents. It has been heavily criticised e.g. by Hector-Neri Castañeda in his review of Hintikka's *Knowledge and Belief*:

> "…[Hintikka's] senses of 'knowledge' and 'belief' are much too strong […] since most people do not know every proposition entailed by what they know; indeed many people do not even understand all deductions from premises they know to be true." (Castañeda 1964: 133–134)[11]

Together with (**K**), (**Nec**) gives us (3.3) and hence closure of knowledge under material implication. This is a weaker principle than (**Nec**), but it may still be deemed unacceptable. The presence of (**Nec**) in a logical system makes it clear that what it models at best a highly idealised notion of knowledge. It has proven very difficult to spell out the degree and the nature of this idealisation in knowledge-based terms.[12]

(**K**), on the other hand, does not demand such a reinterpretation: requiring that the implication has to be known, it does not for itself force upon us an idealisation. It does not, as (**Nec**) does, impute infinitary powers on our agents, nor does it make an unconditional assertion about what every agent knows. Instead, it constrains their behaviour as modelled in the system: (**K**), therefore, is more naturally taken to be an external constraint on our knowledge *ascriptions*: we should not, in other words, deny our agents the capacity to perform *modus ponens*. (**Nec**), on the other hand, forces us to ascribe only deductively closed belief or knowledge sets, i.e. it forces us to ascribe knowledge that normal agents do not have. I conclude, then, that (**K**) may be plausibly held to be acceptable, at least in the absence of (**Nec**) which is the real culprit and a far more problematic principle.

Reflexivity and consistency

Another crucial feature of any acceptable logic of knowledge is the following "truth axiom", incorporating condition (i) in the classical definition of knowledge (**JTB**):

(**T**) $\vdash \mathsf{K}p \to p$

While the inclusion of (**T**) in any logic of knowledge is not controversial, its equivalent for a logic of belief has been subject to disputes. Hintikka, in adapting the logic of knowledge – which he took to be **S4** – to the case of belief replaced (**T**) with (**D**):

(**D**) $\vdash \mathsf{B}p \to \neg\mathsf{B}\neg p$

thereby ruling out inconsistent belief sets.

Lemmon violently criticised this on the ground that (**D**), even if perhaps a useful principle in the representation the agents gives of his own belief,[13] is prohibitively crippling from the modelling perspective – since it is plainly possible (and actual) that people hold inconsistent beliefs. CHECK: WASN'T HE GAY HIMSELF?

3.2.5   Introspection

On the semantic side, an agent's knowledge that *p* is usefully analysed as *p*'s being the case in all of *a*'s epistemic alternatives. The following axiom, sometimes labelled "positive introspection", ensures that the relation of epistemic alternativeness is transitive:

(4) $\vdash \mathsf{K}p \to \mathsf{K}\mathsf{K}p$

---

11. While Hintikka defended (**Nec**) for his notion of knowledge, he grants the point for the – in his view, more general – notion of information: "It simply is not true that there are no objective senses of information in which logical inference can add to our information." (Hintikka 1970: 181)

12. Cf. the following remarks of Hintikka: "What causes the breakdown of these rules [(**??**) and (**??**)] is after all the fact that one cannot usually see all the logical consequences of what one knows and believes. Now it may seem completely impossible to draw a line the implications one sees and those one does not see by means of general logical considerations alone. A genius might readily see quite distant consequences while another man may almost literally 'fail to put two and two together'. Even for one and the same person, the extent to which one follows the logical consequences of what one believes varies with one's mood, training, and degree of concentration." (Hintikka 1970: 181)

13. Though he doubted even that: "Finally, in moments of self-honesty, a man may *admit his own* inconsistency (consistently) in the form of Moore's paradoxical assertion - "my problem is that I'm a homosexual but I don't really believe that I am."" (Lemmon 1965: 382–383)

Jaakko Hintikka argued for (4) on the basis of his peculiar semantics given in terms of defensibility (consistency) of belief sets:

> "That q is the case can be compatible with everything a certain person – let us assume that he is referred to by a – knows only if it cannot be used as an argument to overthrow any true statement of the form "a knows that p." Now this statement can be criticised in two ways. One may either try to show that p is not in fact true or else try to show that the person referred to by a is not in a position or condition to know that it is true. In order to be compatible with everything he knows, q therefore has to be compatible not only with every p which is known to him but also with the truth of all the true statements of the form "a knows that p."" (Hintikka 1962: 18)[14]

The idea, in a nutshell, is that I only know something if it is not excluded by what I know nor by my knowing what I know. This is based on the observation that explicit self-ascription of knowledge is a particularly strong form of endorsing a claim:

> "I am not in a position to say "I know" unless my grounds for saying so are such that they give me the right to disregard any further evidence or information." (Hintikka 1962: 20)

This argument from entitlement or rhetoric force of knowledge claims[15] is inconclusive, however. It may be defended (at least by Hintikka-style arguments) only in contexts where the "absurdity" correlated to an indefensible belief set is understood in performatory terms.[16] This performatory aspect of knowledge claims, however, is cancellable and applies only to first-person utterances. Whenever we are interested in modelling the epistemic behaviour of (real or idealised) agents, such considerations are inapplicable.

In reply, Hintikka relies on the idealisations he made: he only considers statements made on one and the same occasion and presupposes that the person referred to by "a" in the index of the knowledge operator knows that he is being referred to (Hintikka 1962: 106).

Hintikka considers it one of the major advantages of his theory that it is able to explain the pragmatic weirdness of 'Moore's paradox' by its unbelievability:

(3.5) p, but I do not believe that p

Assuming doxastic alternativeness (DA) to be transitive, it is straightforward to show that Moore's paradox is unbelievable.[17] We should, however, resist the temptation to take doxastic alternativeness to be transitive. For it is clearly possible to believe that one lacks beliefs one in fact has. What makes the indefensibility of Moore's sentence paradoxical is not only that we know that, for most if not all p, it might be true and for most of the true p's of our language it is actually true. It also is paradoxical because there does not seem anything wrong with ascribing belief in it *to someone else*. Using transitivity to show that Moore's paradox is unbelievable *tout court* misses this feature: for someone who is claiming that p is true although he does not believe it is claiming that transitivity fails for him. Because he, after all, might very well be right, we get a Moorean paradox at the meta-level.

---

14. He had an ancillary argument for transitivity, based on the observation that "knowing" and "knowing that one knows" are often used interchangeably in ordinary language – or, at least, that their "basic meanings" as captured by an explanatory model are the same (cf. Hintikka 1968: 8).

15. Hintikka sometimes also puts the point in terms of justification: "In the primary sense of *know*, if one knows one *ipso facto* knows that one knows. For exactly the same circumstances would justify one's saying "I know that I know" as would justify one's saying "I know" *simpliciter*." (Hintikka 1962: 28,111)

16. Hintikka is clear on this point: "The absurdity of doxastically indefensible sentences is of *performatory character*; it is due to doing something rather than to the means (to the sentence) which is employed for the purpose." (Hintikka 1962: 77) He calls transitivity also "the quasi-performatory aspect of the verb *know*" (Hintikka 1962: 55). He repeatedly stressed this point in subsequent discussions: "[…] for someone to know that p his evidence […] has […] not only to be good but as good as it […] can be. It has to be such that further inquiry loses its point (in fact, although it is logically possible that such an inquiry might make a difference). The concept of knowledge is in this sense a 'discussion-stopper'. It stops the further questions that otherwise could have been raised without contradicting the speaker." (Hintikka 1968: 13)

17. Assume there is a possible world w where a believes that p and that a does not believe that p. Assume there is a doxastic alternative v for a in w. Then $v \models p \land \neg \mathsf{B}p$. Because of the second conjunct, there is an doxastic alternative u for a in v such that $u \not\models p$. By transitivity, u is also a doxastic alternative for a in w. Hence $u \models p$, which is impossible. So a has no doxastic alternatives in w.

The sense in which (5.1) is indefensible is better brought out by considering that belief in (5.1) makes the world in which it is held either doxastically inaccessible to itself or contradictory. (5.1) cannot be rationally believed in 'reflexive' worlds, i.e. worlds w such that $w\mathsf{DA}w$.[18] Taking a doxastic alternative to be possibly actual as opposed to just merely possible, however, means taking it to be doxastically accessible to itself. So no one can take a doxastic alternative in which (5.1) is true to be a possible way *his* actual world, the world of the believer, might be.

What lies at the bottom of indefensibility of (5.1) is not the transitivity of doxastic alternativeness but the commonly made presupposition that what one says might be true even in a state of complete information. If I utter p, I therefore commit myself to the claim that p might be true even if I knew everything about the actual world there is to know, that is, even if the actual world were my only doxastic alternative.[19] If you utter p, you must consider it possible that you would believe p even if you had a maximally specific belief set, i.e. if you would believe all the truths (or, equivalently, disbelieve all the falsehoods). If the world in which we believe them were our only doxastic alternative, belief in (5.1) would make that world inaccessible to us.

The strongest argument against (4), I think, is the simple fact that knowledge may supervene on factors unknown to the agent. This is nicely brought out in Dretske's notion of a channel condition,[20] but the point can be made independently: Suppose that whether or not I have knowledge of the truth that p depends on the obtaining of some external factors r (i.e. a reliable connection between my belief and the fact that p, the trustworthiness of my conversational partner, the actual absence of other possibly misleading evidence or what have you) such that, if r were not be the case, I would not know that p. It is enough, then, for my knowing that p that r obtains – I do not have to know that it obtains. In order to know that I know that p, however, I have to rule out that $\neg r$, at least according to Nozick's definition of knowledge (**NO**) – but also, if $\neg r$ is sufficiently likely to occur and the demands on justification are sufficiently high, according to the classical definition (**JTB**). To have knowledge that I know that p, I have to rule out more ways the world might be, and thus there are situations in which, lacking this knowledge, (4) is false. An especially vivid case is where I even lack the concepts or linguistic resources to express r: it seems not clear, in this case, how I might be able to rule out that $\neg r$, even while I do not have to articulate the background conditions on which may ordinary, first-level knowledge claims depend.

In the context of epistemic modal logic, (4) has been interpreted as the axiom of *positive introspection*. The following axiom may then be aptly be called *negative introspection*:

(5) $\vdash \neg\mathsf{K}p \rightarrow \mathsf{K}\neg\mathsf{K}p$

Even researches who subscribed to (4) have heavily criticised (5), most notably Wolfgang Lenzen (cf. e.g. Lenzen (1978: 79) and Lenzen (1979: 35)) and Jaakko Hintikka (Hintikka 1962: 106).

(5) ascribes to agents the capacity to gain introspective knowledge just in virtue of lacking another piece of knowledge. While this principle may be acceptable if $\neg\mathsf{K}p$ is given some "positive", "substantial" interpretation (not just lacking knowledge that p but positively disbelieving p or doubting that p), it is inacceptable iff $\neg\mathsf{K}p$ just means that p is false in some of the agent's epistemic alternatives. For such epistemic alternatives are just total states of the world *not excluded* by the epistemic state of an agent at a given time. The agent does not have to be able to survey the range of these alternatives and to know, e.g., whether there is some $\neg p$ world among them: he has, in other words, epistemic access to them only under the description "worlds where all propositions I know are true" – he does not have to know of every proposition whether it is true or false in his epistemic alternatives. Whether or not p fails in some epistemic alternative cannot be decided by the agent on the sole basis of what he knows; he needs, in addition, some grasp of what he could but does not know.

---

18. Suppose $w \models \mathsf{B}(p \wedge \neg\mathsf{B}p)$ and $w\mathsf{DA}w$. Because of $w \models \mathsf{B}\neg\mathsf{B}p$, $w \models \neg\mathsf{B}p$. But the first conjunct becomes $w \models \mathsf{B}p$, so w is contradictory.

19. Something along this line was Hintikka's original justification for the transitivity of his belief relation: "If something is compatible with everything you believe, then it must be possible for this something to turn out to be the case together with everything you believe without making it necessary for you to give up any of your beliefs. If your beliefs are to be consistent, it must also be possible for all your beliefs to turn out to be true without forcing you to give up any of them." (Hintikka 1962: 24) Williamson takes this robustness of knowledge to be one of its distinctive differences from true belief: "Knowledge is superior to mere true belief because, being more robust in the face of new evidence, it better facilitates action at a temporal distance." (Williamson 2000: 101).

20. Dretske even denies (4) even for the case where the first modal operator is weakened to mere belief: "We naturally expect of one who knows that P that he believes that he knows, just as we expect of someone who is riding a bicycle that he believe he is riding one, but in neither case is the belief a *necessary* accompaniment." (Dretske 1971: 21)

The acceptance of (**5**) in our epistemic logic forces us to ascribe knowledge to epistemic agents on the sole basis of their *lack* of knowledge of some other proposition. Whenever someone does not know that p, e.g. because he has never heard of p or because he even lacks the concepts to formulate a sentence or to entertain a thought expressing it, (**5**) (together with Modus Ponens), forces us to ascribe him knowledge of his lack of knowledge. The world would be very different if people were always that knowledgeable about their ignorance. Moreover, (**5**) meshes badly with the confidence we naturally have in our knowledge claims. Suppose I am subjectively certain that a false proposition p is true. Given (**T**), I do not know it. (**5**) then ascribes to me knowledge that I do not know that p, which (given that I proportion my degree of confidence in p to my knowledge) seems to rule out my being subjectively certain that p. Unless "knowledge" is understood in an idealised sense, (**5**) seems to me out of question.[21]

Given the six principles above, it is now possible to define different modal logics. The weakest normal modal logic **K** is given by any complete axiomatisation of the propositional calculus, (P), (**K**) and the necessitation rule (**Nec**). We add (**D**) or (**T**) to get **D** and the stronger **T** respectively, and add (4), (**5**) and (**D**) to get **K45**, **KD45**. Adding (4) to **T** gives us a system called **S4**, while adding (**5**) to **T** gives us **S5**.

3.3   The A Priori

3.3.1   The Traditional Notion and its Problems

Something like the contemporary notion of the a priori surfaces (perhaps for the first time) In Descartes' *Meditations*, where he announces, after and as an upshot of the argument from (something like) the argument from illusion, never to trust anything he has came to believe on the strength of testimony received from and through the senses:

Nempe quidquid hactenus ut maxime verum admisi, vel a sensibus, vel per sensus accepi; hos autem interdum fallere deprehendi, ac prudentiae est nunquam illis plane confidere qui nos vel semel deceperunt. (AT VII 18¹⁵⁻¹⁸)	Tout ce que j'ay receu iusqu'à present pour le plus vray & assuré, je l'ay appris des sens, ou par les sens: or j'ay quelquefois éprouvé que ces sens estaient trompeurs, & il est de la prudence de ne se fier jamais entierement à ceux qui nous ont une fois trompez. (AT IX/I 14)

As the continuation makes clear, "per sensus" includes knowledge received through testimony, and based, whether exclusively or just in part, on testimony. The conjunctive criterion thus includes all knowledge about the empirical world, though it concerns it individually – and not collectively as does the dream argument (which, contrary to what is sometimes claimes, *is* the claim that I might always be sleeping).

The "per sensus" criterion, including (putative) knowledge through testimony, covers empirical, but factual knowledge such as my knowledge that Addis Abbeba is the capital of Ethiopia, but it also presumably covers too much, including all the knowledge I have acquired in school.

"I understand 'a priori' to apply to a person's knowledge when that knowledge is underwritten by an apriori justification or entitlement that needs no further justification or entitlement to make it knowledge. A justification or entitlement is *a priori* if its justificational force is in no way constituted or enhanced by reference to or reliance on the specifics of some range of sense experiences or perceptual beliefs." (Burge 1993: 230)

Already in its first use, however, the 'a priori' / 'a posteriori' distinction is not just about the – actual or possible – source of the belief the so qualified knowledge claim claims to constitute knowledge. It is about justification, or rather certain types of justification, supposedly one-to-one correlated with the types of contents the beliefs in question have. We thus have, as a first approximation

(**AP**)   A belief that p is a posteriori iff it, directly or indirectly, has to be justified with recourse to sense experience.

Identifying the a priori with what is not a posteriori, we get the familiar notion: a belief is a priori iff it is possible to justify it without recourse to sense experience. Obviously, the definition is problematic in many ways: what

---

21. It may thus be argued that (**5**) forces us to ignore the possibility of error, i.e. to assume that agents are always right in their knowledge claims: this, as Stalnaker (1999: 258) noticed, blurs the distinction between knowledge and belief.

counts as 'recourse', what as 'sense experience', what as 'possible', what as 'justify'? Will a belief in a theorem count as a posteriori iff all possible proofs of it are so complex that they can only be checked using a computer? Will Descartes' certainty of his existence, arrived at through introspective reflection on what he is doing when he doubts it, count as a posteriori? Will count as justified claims that may be argued to be probable from the armchair, as e.g. default assumptions in statistics?

Of all concepts in (**AP**), possibility is perhaps the most problematic: to extend it beyond the presently feasible, but not as far as what is metaphysically possible, appeal is made to the notion of an "ideal thinker", so as to expand the purvey from what (type of) justification our beliefs have or could have to what (type of) justification they (epistemically) should have: a priori beliefs are those for which an ideal warrant (i.e. the warrant that a being without limitations of computational capacities, i.e. an ideal thinker, has) is immune to empirical augmentation or diminishment (or defeat).

It is not clear, however, what "without limitations of computational capacities", and thus the idealisation in question, involves. There is, of course, a worry of circularity: How to spell out in what respect the ideal thinker is different from us and similar (it has to be similar for if it is too different, it can know a priori things that we would like to count as a posteriori knowable) to us without using the idea of the a priori (or empirical indefeasibility by empirical evidence etc.) It would not do, e.g., to define a priori beliefs as those the ideal thinker could come to have without relying on experience and then defining an ideal thinker as one who grasps all a priori truths. What the ideal thinker can know a priori is what *we* could know without empirical evidence had we unlimited computational capacities (explanation of what it can know).

More pressing even, there is a problem about "computational". Take Kant. For Kant, an ideal thinker, to whom everything would appear as it is itself just by the fact that it so appears (ie. someone who would need to be a transcendental realist, not in need of the appearance / in-itself distinction), is someone who has an "intuitive intellect". Though the details are murky, it is clear that one dimension of the contrast between us, who have a discursive intellect, and such an ideal thinker is that we can synthetise (roughly: form representations by grouping features) only sequentially, while an ideal thinker intuits the manifold of the given all at once – a difference which, for him, clearly has to do with our computational powers being qualitatively, not just quantitatively limited.

An alternative would be to take "a priori" / "a posteriori" classifies (the warrant for) *actually possessed* items of knowledge. It does not make sense to ask of a truth "p" whether it would be known a priori or not, we have to add *by whom* it would be known in this or that way. At least for truths "p" that are actually known by at least someone, however, we seem to have a clear-cut distinction:

(3.6) p is a priori $:\Longleftrightarrow$ $\exists x(x$ knows that $p \wedge x$'s warrant could not be

defeated by empirical considerations)

We therefore allow for a posteriori warrant, as long as it is purely incremental: we require that the x's belief that p would still be justified even if he lacked any empirical warrant for his belief that p.

Burge 1993: Some beliefs with a priori justifications or entitlements are vulnerable to empirical overthrow.

Williamson, Philosophy of Philosophy, says similar things

Another possibility is to treat "a priori" as an evaluative notion. What we justifiably treat as a priori, for an evaluationist, *is* a priori. Field (2000: 117) defines an a priori proposition as one that can be reasonably believed without empirical evidence (is 'weakly a priori') and that admits of no empirical evidence against it (is 'empirically indefeasible'), i.e. empirical evidence which is such that it lowers its ideal credibility. Default reasonable propositions (propositions which can be reasonably believed without any justification at all) are a priori iff they are empirically indefeasible (Field 2000: 119). Reasonableness is taken to be an evaluative property in the sense of Gibbard (1990) and Field (1994):

> In calling a rule reasonable we are evaluating it, and all that makes sense to ask about is what we value.
> (Field 2000: 138)

3.3.2 The Extent and Nature of the A Priori

In addition to the problem of definition, there are at least three other families of philosophical questions about the a priori: its source, its nature and its extent. Regarding the problem of source, it has seemed mysterious to many how, i.e. by the exercise of what faculty, beliefs that are not justified 'by' the senses, can be justified at all. Do we have to accept some form of pure intellect, a faculty of rational intuition, intellectual 'perception' of mathematical and conceptual truths?

Discussions of the nature of the a priori have centered around the notion of proof.

Explain why this is remarkable: What we know a priori, by any reasonable definition of the latter notion, is something we do not know by or on the strength of perception or perceptual beliefs. But whereby do we know it?

What kind of justification do we acquire by constructing, understanding or sequentially following a mathematical proof?

Most discussions, however, have been of the extent of a priori knowledge. While traditionally mathematical knowledge had been taken to be the paradigm case of the a priori, a distinction introduced by Kripke has been proven to be a rich source of claims to a priori knowledge. While discussion the question whether proper names have Fregean senses, i.e. whether they semantically contribute more information than is given by their primary function of making their referents available as subjects of predication, Kripke distinguishes two uses of definite descriptions such as "the teacher of Alexander" to pick out individuals. When we use them to *give the meaning*, we instruct our interlocutors to pick out by them (and understand us as picking out) whoever satisfies the description. When we use them to *fix the reference* of the term, however, we rely on a (presumably shared) understanding that we use them to talk about *that* individual, whether or not she in the circumstances described really was teaching Alexander:

> Following Donnellan's remarks on definite descriptions, we should add that in some cases, an object may be identified, and the reference of a name fixed, using a description which may turn out to be false of its object. […] In such cases, the description which fixes the reference clearly is in no sense known *a priori* to hold of the object, though a more cautious substitute may be. If such a more cautious substitute is available, it is really the substitute which fixes the reference in the sense intended in the text. (Kripke 1980: 80, n. 34)

It is in this referential use that we fixed what counts as a meter by laying down the convention that being one meter long *is* having the length (whatever it is) of the standard metre in Paris. Our adoption of this convention, according to Kripke, gives us a priori knowledge: If we use stick S to fix the metric system, then we know "automatically, without further investigation, that S is one meter long" (Kripke 1980: 56). This is surprising because the standard metre itself is only *contingently* of the length it actually has:

> If someone fixes a meter as 'the length of stick S at t_0', then in some sense he knows *a priori* that the length of stick S at t_0 is one meter, even though he uses this statement to express a contingent truth. But, merely by fixing a system of measurement, has he thereby *learned* some (contingent) *information* about the world, some new *fact* that he did not know before? It seems plausible that in some sense he did not, even though it is undeniably a contingent fact that S is one meter long. (Kripke 1980: 63)

Suppose we introduce the name "Julius" by stipulation for the inventor of the zip, whoever that is. It is then supposed to be a priori that Julius invented the zip, even though this latter truth does *not* have the same logical form as "The inventor of the zip invented the zip".

3.3.3 Two-dimensional Semantics and Semantic Externalism

Intro

Crucial problem: if we bestow our words or concepts with the power of making distinctions *we* cannot make, we put some distiance between us and our words; the danger is overgeneralisation.

3.3.4 The McKinsey Paradox

The so-called "McKinsey's paradox" is the following:

(3.7)   major   *a* is a priori justified in believing that (if *a* believes that dogs either bark or don't bark, then dogs have existed)

(3.8)   minor   *a* is a priori justified in believing that *a* believes that dogs either bark or don't bark.

(3.9)   closure   For any *p*, *q*, if *a* is a priori justified in believing that *p* and $p \rightarrow q$), (then *a* is a priori justified in believing that *q*.

(3.10) paradox   *a* is a priori justified in believing that dogs have existed.

Schiffer's 'solution' proceeds from the following definition: A belief is *strongly a priori justified* if it is not a posteriori justified and it's warrant does not include the fact that one is a posteriori justified in holding a certain other belief. A belief is *weakly a priori justified* if it is not a posteriori justified but it's warrant includes the fact that one is a posteriori justified in holding a certain other belief.

Deny closure (3.9) for weakly a priori justified beliefs and show that *a* is only weakly a priori justified in believing that if *a* believes that dogs either bark or don't bark then dogs have existed (3.7).

Motivating example: *a*'s warrant for a prima facie a priori belief may be increased by a posteriori considerations:

W1 establishment of a proof of mathematical theorem T by calculation

W2 assent to the proof by other mathematicians

together form *a*'s warrant for believing T. W2 indisputably increases the warrant for his belief, but is a posteriori.

3.3.5 Inferentialism and McGee

Paul Boghossian has proposed both a theory of understanding and a theory of meaning for expressions usually called 'logical constants'. Within the theory of meaning, he defends *implicit definition*: arbitrary stipulation of the truth of the familiar introduction and elimination rules give the logical constants their meaning, i.e. determine their meaning to be the truth function which make the introduction and elimination rules truth-preserving. Within the theory of understanding, he holds that every competent speaker has implicit knowledge of the theory of meaning, i.e. is disposed to infer according to the introduction and elimination rules.

With respect to the conditional (what is expressed by "if …then $\cdots$" and the rule of inference

$$\frac{p \quad p \xrightarrow{} q}{q} \textbf{ MPP}$$

we can distinguish two claims:

- that acceptance of the validity of MPP is necessary for competence with "if …then $\cdots$"

- that MPP implicitly defines (or helps to define) the meaning of "if …then $\cdots$"

The combination of these two theses allows him to explain why blind rule-following can be justified, i.e. why someone can be cognitively blameless who reasons in accordance with a rule he is unable to cite the rule and to produce a justification for it. Boghossian's answer is that we are entitled to follow MPP blindly because being disposed to infer according to MPP is a precondition for having one of the concepts ingredient in it, i.e. the concept expressed by "if …then $\cdots$". MPP itself can then be justified by rule-circular justification, i.e. by applying MPP to the truth-table of "if …then $\cdots$". We are entitled to this use of MPP while justifying MPP because we could not have the concept of "if …then $\cdots$" in the first place would we not be (prepared to) accept MPP.

Another advantage of Boghossian's theories of both the meaning and the understanding of logical constants is that they align themselves neatly with some familiar claims about singular terms and the two-dimensional semantics of natural kind terms: according to the two-dimensional picture of the semantics of "water", its meaning is what actually fits the reference-fixing description while competent speakers are required to have implicit knowledge of the rule thereby employed – they are to know not just what the meaning is but how the meaning has been fixed and would have been fixed in different ways under different circumstances. Similarly, while the meaning, or content, of "I" is just its referent, a competent speaker is expected to have the disposition to draw the inference from "I am in Buenos Aires" to "the speaker is in Buenos Aires". More generally, one often finds the combination of a causal theory of reference within the theory of meaning combined with a requirement of implicit knowledge of that theory of meaning within the theory of understanding.

Attractive as this general picture is, we think that it meets with a difficult dilemma. If it is going to provide us with a workable notion of the a priori that is in touch with traditional uses of this notion, "acceptance of the validity of MPP" has to mean more than just reasoning in accordance of it or acceptance of its use: the same is true of obvious truths and communally shared presuppositions; most people, after all, unreflectively assent to trivialities such as "grass is green" and "dogs are animals". If acceptance of the validity of MPP, however, means more than just unreflective and underived assent, the theory of understanding is going to declare incompetent speakers of the language who hold heretic or unusual beliefs *about* the logical constants.

The theory of understanding of "if …then ⋯" will have to deal with what one may call the "McGee problem": someone who disputes or even only doubts the validity of MPP has to be treated as misunderstanding "if …then ⋯". The eminent philosopher-logician Vann McGee has written a paper entitled "A Counterexample to Modus Ponens", claiming that MPP has, at best, restricted validity (McGee 1985). Other logicians and philosophers have discussed for years the validity of disjunctive syllogism. The problem here is not only that, at least among polite persons, doubting the very intelligibility of your opponents' claim should be the very last resort in the dialectic. It also seems possible to turn the McGee- into a Pierre-case. McGee asserts, in his language, that MPP is not valid, while he need not dispute, indeed cannot dispute, MPP as it is used by Boghossian. Should we now say that McGee believes or that he does not believe that MPP is valid? Another problem for Boghossian's theory of understanding "if …then ⋯" concerns future and past logicians: any future discovery that MPP is valid for only a restricted range of propositions would imply a radical concept-change and bring with it the same kind of incommensurability problems that long have plagued theorists of scientific theory change.

Problem: insofar as this presupposes a *special* notion of concept mastery, it may be as circular as the other notions previously discussed.

– THIS MAY SERVE ALREADY TO INTRODUCE THE FAILURE OF WARRANT TRANSMISSION NOTION

Chapter 4

Skepticism

4.1 Different Questions Need Different Answers

cf talk by Manuel Pérez Otero, Dimensiones del escepticismo epistemológico, perhaps this is now published as Pérez-Otero (2011) (file)

4.1.1 Pyrrhonian and Modern Skepticism

4.1.2 Different Skeptical Questions

The general form of the sceptical argument is:

$$\{A_1, \ldots, A_n\} \models \neg W_{x,t} p$$

where W ("warrant") is any relationship that mandates acceptance of a proposition (in virtue of which the thinker is in a position to accept a certain proposition): knowledge, justification, belief (in some sense). Modality comes in only at the level of the specification of a particular reading of W and typically falls short of logical necessity.

Whether or not someone possesses warrant for a certain proposition is a normative issue:

> "Acceptance-support is a normative notion: a relationship between a thinker and a proposition is acceptance-supporting if it suffices to warrant any thinker's acceptance of the proposition in question."

Warrant may be strong or weak:

- x's warrant at t for p is *weak* iff "$\neg W_{x,t} p$" is consistent with x's standing in some other acceptance-supporting relationship to p;
- otherwise, x's warrant at t for p is *strong*.

In view of Russell's retreat – to give up on knowledge and to retreat to some other notion of rational acceptance –, our response to the sceptic questioning possession of a certain kind of warrant for some proposition, should work equally for stronger kinds of warrants if the sceptic argument could easily be rephrased in terms of some such stronger kind of warrant.

According to Williamson, knowledge is strong, for it not only mandates assertion but also belief.

In order to defeat scepticism, the knowledge secured against the sceptic should be *claimable*; this means that second-order scepticism is equally to be avoided, the general form of argument of which is the following:

$$\{A_1, \ldots, A_n\} \models \neg W_{x,t} (W_{x,t}) p$$

The second-order warrant may consist in something different from what constitutes the first-order warrant. Typically, it is an internal kind of warrant.

Some warrant is *internal* if "$W_{x,t}p$", if true, may be certified by deploying just resources of ordinary self-knowledge, reasoning and a priori reflection. These are just the resources you deploy to justify a claim that you acted for the best – reasoning on your motives, reconstructing (at best) the practical syllogism you acted on and reflecting on your good conscience. An (internally) justified belief is a belief that stands up to this kind of scrutiny. Some warrant is *external* if the truth of "$W_{x,t}p$" involves relations between x and the world which are not so certifiable.

Cartesian sceptical arguments work by floating some claimed undiscountable Horrible Possibility which does away with a whole tract of what we take to be reality or dislocates us from it. *Humean* sceptical arguments work by claiming a vicious circle in our justificatory procedures in an area which undermines our justification or knowledge for it by challenging our warrant for ampliative inferences of some kind. Humean scepticism with respect to induction is often presented as pointing to the fact that inductive inferences are deductively valid only if supplemented with a further tacit premiss which in turn can only be justified inductively. Humean scepticism, however, targets any kind of ampliative inference and is not to be remedied by an appeal to uniformity of nature (for, even given this extra premiss, the inductive argument is still deductively invalid, at least if we do not know that the observed regularity qualifies as a law of nature). Some inferences are *essentially* ampliative, in the sense that additional information is required to make them rational (not: to make them deductively valid or to force the transition). The problem then arises if the required additional information is only justifiable by the same kind of ampliative inference.

A *direct* sceptical argument is one which exploits just the purported unwarrantedness of "$\neg H$" where "H" is some large sceptical possibility, and closure of W, to unsaddle W for every p such that $p \to \neg H$. A classic case is the brain-in-a-vat scenario. Apart from closure, it is also presupposed that the mere fact that you cannot rule out "$\neg H$" is enough to establish "$\neg W \neg H$":

$$(4.1) \quad \frac{\begin{array}{c} p \to \neg H \\ \neg W \neg H \end{array}}{\neg W}$$

Direct arguments are difficult to construct: you have to find a coherent scenario which is inconsistent with the truths of a large number of beliefs while still being consistent with them having the content they have.

An *indirect* sceptical argument works by exploiting the inconsistency of "H" with the normal cognitive pedigree of the range of beliefs covered by "p", rather than their truth. The Dreaming argument is a classic case (if I am right now dreaming, I could not have the warrant I claim to have that it is now raining, even if my dream is in fact caused by the thunderstorm – the required *systematic* relationship normally present in perception is missing):

$$(4.2) \quad \frac{\begin{array}{c} Wp \to \neg H \\ \neg W \neg H \\ \hline \neg W \neg \neg Wp \end{array}}{\neg WWp}$$

If W is iterative (if we have $Wp \models WWp$), the indirect sceptical argument collapses into the direct one.

4.1.3 Different Anti-Skeptical Answers

With Pryor, we can distinguish between two types of answers to the skeptic:

1. ambitious anti-sceptical project: refuting the skeptic on his own terms.
2. modest anti-sceptical project: establishing "tou our satisfaction that we can justifiably believe and know such things as that there is a hand without contradicting obvious facts about perception".

4.1.4 Burge on Content Preservation

(Burge 1993)

Burge on content preservation: (Burge 1993)

– Against Chisholm (1977): Reliability of memory in its function of content preservation is a background condition, "[b]ut this preservation is not part of the justification of the explanation, nor does it add to it – even though if it were to fail, the explanation would be jeopardized." (Burge 1993: 234) – Against Chisholm (1977); Ross (1975): Similarly for testimony. Beliefs are not empirical just in virtue of being acquired from others. – Based on "one's reason transpersonal function of presenting the truth, independently of special personal interests." (Burge 1993: 243)

4.1.5   Undermining Truth or Claimability? Moore's Proof

Lewis (1996), Wright (2004)

(Pryor 2000)

Iterability and containment: Is knowledge that I know always the result of / analysable as iteration? Not sure. Compare: doubt.

4.2   Humean Skepticism

Hawke (2017), which is in Fischer & Leon (2017b) (file), which also contains (perhaps to be put into the essence-readinglist)

1. Fischer & Leon (2017a)
2. Nolan (2017)
3. Tahko (2017)
4. Shalkowski (2017)
5. Bueno (2017)
6. Vaidya (2017)
7. Rasmussen (2017)
8. Kung (2017)
9. Biggs (2017)
10. Krödel (2017)
11. Hanrahan (2017)
12. Roca-Royes (2017)
13. Leon (2017)
14. Fischer (2017)

4.2.1   Warrant Transmission Failure, Rule Circularity and Blind Rule-Following

Wright's template An argument of the form

$$(4.3) \quad \frac{p \quad p \xrightarrow{p} q}{q}$$

fails to transmit warrant iff there is a proposition r such that

- the thinker's warrant for p consists in his being in a state that is subjectively indistinguishable from some state he could be in if r were false and

- I r were false, then q would be false

In typical cases, p is a proposition about my evidence, e.g. my subjective experience (a type I proposition), q is a so-called "type II proposition", i.e. a truth I take to be justified on the basis of my subjective proposition and r is some so-called "hinge proposition", i.e. a truth such as that there are other minds or that the world did not come into existence five minutes ago.

Wright's diagnosis of warrant transmission failure is then the following: "*p*" is only evidence for "*q*" if (and hence the second premiss is only justified to the extent that) we are justified in taking "*r*" to be true. So "*p*" justifies "*q*" only to the extent it also justifies "*r*". But it cannot justify "*r*", for it is compatible with "¬*r*". Hence it cannot justify "*q*".

4.2.2 The Justification of Deduction

Suppose I know that *p* and that if *p* then *q*. If I do not know *q*, then, for all I know, ¬*q* might be true. This means that ¬*q* is consistent with everything I know. So *p* cannot be something I know, for ¬*q* is inconsistent with *p*. This means that someone who denies closure has to argue against one of the following:

- If I don't know that *q*, then for all I know, ¬*q* might be true.
- If, for all I know, ¬*q* might be true, then ¬*q* is consistent with all I know.

Another popular argument for closure is the following: If closure fails, how can reasoning be a means of discovery? how can we obtain knowledge in mathematics? Failure of closure better had to be local. That there are exceptions to closure is consistent with it's being a method of gaining knowledge of conclusions.[1]

To uphold closure because reasoning should be a method of producing knowledge, however, is to confuse closure with transmission. And we know that there are exceptions to transmission. There are differences, however: to have transmission across entailment, we have to carry out the proof of *q* from *p* − what we presumable are not required to do for closure.

Dretske is not really giving a counterexample to closure, however, but a case of transmission failure: Visual perception is only evidence for the premiss if the conclusion is already taken for granted. The route to warrant the premiss goes through the conclusion. Dretske gave an argument against "$K^*p \wedge K(p \Rightarrow q) \models K^*p$", where K^* is a particular way of coming to know a certain proposition.

What you have the permission to take for granted cannot be supported by something which is evidence for a premiss from which it follows only if it *is* taken for granted.

Tracking conditionals are only appropriate if the operator is factive. They are not appropriate, e.g., for all-things-considered justification. If your attitude towards "*p*" is not factive, the nearest world in which *p* is false and you still have the belief may be the actual world.

Closure holds but transmission fails: the warrant for the premiss does not transmit to the conclusion but I still know the conclusion by some other means than the means by which I know the premiss.

4.2.3 Rejecting Empiricism?

Some examples for what Wright calls "I/II/III arguments":

matter	mind	past
My current experience is . in all respects as if *p*	For some other person *a* distinct from myself: *a*'s behaviour and physical condition are in all respects as if *a* was in mental state *m*.	It seems to me that I remember it being the case that *p* yesterday.
p	*a* is in mental state *m*.	It was the case that *p* yesterday.
There is a material world.	There are minds besides my own.	The world did not come into being today replete with apparent traces of a more extended history.

GENERAL PROBLEM: p is justified; 2 is ok; BUT (3) only if (1) is ok, which it is, but only on the presupposition of 3.

The I/II/III paradox:

Cognitive locality: (In this area,) type-II propositions can only be known on the evidence of (by ampliative inference from) type-I propositions.

---

1. There are other arguments against closure: given closure, every necessary truth is in principle knowable; it is not very clear how closure might deal with undecidable propositions.

Ampliativity: The evidence provided by type-I propositions for type-II propositions is information-dependent, requiring (among other things) collateral warrant for a type-III proposition.

So: Knowledge of type-III propositions cannot be achieved by transmission of evidence provided by type-I propositions for type-II propositions across a type-II to type-III entailment – rather it's only if one already has knowledge of the type-III proposition that any type-II proposition can be known in the first place.

However,

Empiricism: Type-III propositions cannot be known in any other way.

The general principle is

No bootstrapping: In a case where H is probabilified by e only under assumptions (collateral information) $\{I_1, \ldots, I_n\}$, the warrant constituted by e for H does not transmit to I_i.

There is a difference between

- e is evidence for H only given $\{I_1, \ldots I_n\}$
- S is justified in using e as evidence for H only if $\phi(\{I_1, \ldots I_n\})$, where "$\phi$" stands either for the truth predicate or for some relation between S and $\{I_1, \ldots I_n\}$.

Does not the first reduce to the second if the relevant notion of evidence is closed under implication (which it should be)? The Zebra case is only a failure of closure-under-visual-evidence: the conclusion you have for them not being cleverly disguised mules is partly inferential. To answer the sceptic, we have to answer the second worry anyway.

- the dogmatist response: the attitude we are required to have towards the type-III proposition for the inference to be warrant-transmitting it that we are not doubting its truth. But this falls foul on our intuitions; being open-minded is not enough BUT not doubting does not entail being open-minded.
- the externalist response: the type-III propositions that are part of the collateral information needed for the type-I propositions being evidence for some type-II proposition are only required to be true but neither required to be believed nor to be warranted. But we still need a transition from truth to warrant!

Crispin Wright rejects Empiricism: Wittgensteinian Certainty and Kantian conditions of possibility

But the symmetry case has not been made! Descartes at least tries

4.2.4 Dogmatism and Perceptual Justification

4.3 Cartesian Skepticism

4.3.1 Cartesian Doubt: Methodic and Hyperbolic

this is in necessary-truths.tex

4.3.2 Doubting the Material World

this is in necessary-truths.tex

4.3.3 The Brain in the Vat

Hilary Putnam hat im ersten Kapitel seines Buchs *Reason, Truth and History* (1981) eine Hypothese aufgestellt, die oft in engen Zusammenhang mit Descartes' Täuschergotthypothese gesetzt, von manchen sogar mit ihr identifiziert wird.[2] Im folgenden möchte ich diese sog. "'brains in a vat'"-Hypothese zuerst darstellen, die Gründe erläutern, die Putnam motiviert haben, sie als inkohärent zu verwerfen, dies kritisieren und schliesslich zeigen,

---

2. Explizite Identifikationen finden sich bspw. bei Massimo Dell'Utri (1990: 79), bei Alex Burri (1994: 125) und bei Hans Poser (2003: 63). Zumindest stark suggeriert wird die Identifikation bei David Davies (1995: 203), Anthony L. Brueckner (1986: 148) und John S. Carriero (1987: 2, 15).

dass die Identifikation mit der Täuschergotthypothese ungerechtfertigt ist. Das von Putnam supponierte und in der zeitgenössischen Skeptizismus-Debatte unter der " Gehirn im Tank"-Etikette diskutierte skeptische Szenario entspricht nicht der Täuschergotthypothese, sondern dem Schlafargument, und es teilt dessen Voraussetzungen, die seine Reichweite ganz wesentlich eingrenzen. Eine eingehende Diskussion wird uns deshalb wesentliche Aufschlüsse über den Unterschied der beiden wesentlichen Zweifelsargumente Descartes' geben.

Putnam diskutiert die Möglichkeit, dass wir Hirne in einem Tank sind, der uns von der Aussenwelt kausal isoliert, während unsere Wahrnehmungen durch ein Computerprogramm produziert werden. Er verwirft dieses Szenario aufgrund der Beobachtung, dass in diesem Fall der Satz "'wir sind Hirne im Tank'" falsch wäre. Putnams Argument macht Voraussetzungen, die zu denen von Descartes' Schlafargument analog sind: Ein Hirn im Tank kann mit seinem Wort "'Hirn'" nicht etwas bezeichnen, zu dem es in keinem kausalen Kontakt steht.[3] Das, was es stattdessen bezeichnet, nennt Putnam "'brain in the image'":

> It follows [von der Voraussetzung, dass Hirne im Tank sich nicht auf etwas beziehen können, mit dem sie in keinem kausalen Kontakt stehen] that if their " possible world" is really the actual one, and we are really the brains in a vat, then what we now mean by "'we are brains in a vat'" is that *we are brains in a vat in the image* or something of that kind (if we mean anything at all). But part of the hypothesis that we are brains in a vat is that we aren't brains in a vat in the image (i.e. what we are " hallucinating' isn't that we are brains in a vat). So, if we are brains in a vat, then the sentence "'We are brains in a vat'" says something false (if it says anything). In short, if we are brains in a vat, then "'We are brains in a vat'" is false. So it is (necessarily) false. (Putnam 1981: 15)

Jemanden, der die Hypothese, wir seien Hirne im Tank, als kohärent ausgibt, bezeichnet Putnam als metaphysischen Realisten. Sein Argument gegen den metaphysischen Realismus lässt sich wie folgt rekonstruieren:

F1 Zwei Dinge können physikalisch identisch sein und sich in ihren semantischen Eigenschaften unterscheiden.

F2 Zwei Dinge können qualitativ (für uns) ununterscheidbar sein und sich in ihren semantischen Eigenschaften unterscheiden.

F3 Zwei qualitativ ununterscheidbare Zeichen können gleich angewendet werden und sich in ihren semantischen Eigenschaften unterscheiden.

F4 Wenn zwei Zeichen dieselben semantischen Eigenschaften haben, müssen sie nicht nur gleich angewendet werden, sondern ihre Anwendungen müssen dieselbe Ursache haben.

F5 Aussagen eines Hirns im Tank haben eine andere Ursache als qualitativ ununterscheidbare und gleich verwendete Aussagen, die wir machen.

F6 Die Aussage "'Ich bin ein Hirn im Tank'", die ich machen würde, wenn ich ein Hirn im Tank wäre, hat andere semantische Eigenschaften als meine Aussage "'Ich bin ein Hirn im Tank'".

F7 Meine Aussage "'Ich bin ein Hirn im Tank'" ist genau dann wahr, wenn ich ein Hirn im Tank bin.

F8 Die Aussage "'Ich bin ein Hirn im Tank'", die ich machen würde, wenn ich ein Hirn im Tank wäre, ist nicht wahr.

F9 Wenn wir Hirne im Tank wären, könnten wir nicht behaupten, wir seien Hirne im Tank.

F10 Wenn wir uns vorstellen können, wir seien Hirne im Tank, können wir nicht behaupten, wir seien Hirne im Tank.

F11 Also ist die Hirn-im-Tank-Hypothese inkohärent.

F12 Also ist es notwendig, dass wir keine Hirne im Tank sind.

F13 Wenn der metaphysische Realist recht hätte, dann könnten wir Hirne im Tank sein.

F14 Also hat der metaphysische Realist unrecht.

2 Interpretations: – to show that we are not BIVs (that's just an application of McKinsey) – to show that metaphysical realism si false because it committed to the *coherence* of the BIV hypothesis

Version of Müller (2001: 302):

---

3. Ein wichtiger Unterschied liegt darin, dass Descartes' Voraussetzung die Bedingungen von Wissensansprüchen betrifft, m.a.W. erkenntnistheoretisch, nicht semantisch motiviert ist.

1. In the language I am actually speaking right now, the word 'tiger' refers to tigers.
2. In the language of an envatted brain, the word 'tiger' does not refer to tigers.
3. Therefore, the language which I am actually speaking right now is different from the language of any envatted brain. (From (1) and (2)).
4. Therefore, I am not a brain in a vat. (From (3)).

An important implicit premise is that "the language I am actually speaking" is a definite description, and picks out just one language. Another one is that words may be identical across languages, an assumption which is itself in tension with externalism (Müller's name for premiss (2)). Thirdly, and most importantly, it is assumed, in the application of PII to get the conclusion, that "x speaks a language in which the word 'tiger' refers to tigers" is an extensional context. To get a valid argument from the template, much more needs to be said on languages. We only get (3) from (1) and (2), if we assume that there is only one language I am actually speaking right now and that all envatted brain speak just one common language, and no envatted brain speaks any other language. We only get (4) from (3) if we assume that I do not speak, and could not learn, the common language of all brains in a vat. Not even this will be sufficient: it has to be assumed that nothing that is not a brain in a vat could speak the brain-in-a-vat language, i.e. that only brains in a vat speak a language where 'tiger' does not refer to tigers. Suppose that I do speak a language, call it Tinglish, that contains the word 'tiger' but where this word does not refer to tigers. Then we get a version of (2) that sounds true: "In the language of a speaker of Tinglish, the word 'tiger' does not refer to tigers", but (4) does not then follow from (3).

Even if the argument were valid, however, it would still not be sound: despite of what Müller (and Putnam) say, premiss (2) is not warranted by externalism. Externalism warrants us, or so let us assume, in denying that the analogue of (1) is true of brains in vats, i.e. in believing that it is not the case that in the BIV's language the word 'tiger' refers to tigers. But again this constitutes a difference between the language of the argument and any language a brain in the vat can speak only if "x contains the word "tiger" and this word in x refers to tigers" is an extensional context.

Stripped from these largely irrelevant complexities, the argument under consideration strips down to:

1. I can talk about tigers.
2. No brain in a vat can talk about tigers.
3. Therefore, I am not a brain in a vat.

This argument is valid, but is it sound? We should not infer (1) just from the presence of the word "tiger" in our language – it is, or at least may be, present in the brains in a vat language as well.

Also, the conclusion is too weak: the reason that we are not brains in a vat may simply be that we are not brains; but we might still be *realised* by brains in vats.

Obwohl es sich begründen lässt, warum wir annehmen, dass wir wach und keine Hirne im Tank sind, lässt es sich meiner Ansicht nach nicht *zeigen* – zumindest nicht, ohne an irgendeiner Stelle der Argumentation vorausset-zen, dass irgendeiner irgendeinmal wach oder kein Hirn im Tank ist, war oder sein wird. Auch wenn wir diese kontingente Voraussetzung machen *müssen*, können wir uns dennoch eine Situation vorstellen, in der wir sie nicht machen *dürften* – und zwar deshalb, weil sie falsch ist. Die Beschreibung dieser Situation entzieht sich unserem Ver-ständnis: wir können sie weder ausschliessen noch für möglich halten. Wie gehen wir mit einer solchen Sachlage um? Descartes würde vorschlagen, dass wir auf ein Wesen rekurrieren, das nicht nur einen weiteren Wirklichkeits-sondern auch einen weiteren Möglichkeitshorizont hat, ein Wesen, für das es nichts Unmögliches gibt. Weil wir die Situation, die wir nicht für möglich halten können, ebensowenig auszuschliessen in der Lage sind, müssen wir einem solchen Wesen die Macht zusprechen, sie herbeizuführen – selbst wenn für uns in diesem Fall etwas wirklich wäre, was wir nicht einmal für *möglich* halten können. Dies ist die Gefahr, auf die der Skeptiker Nagels aufmerksam macht: dass wir uns in einer Weise täuschen könnten, in der wir uns eine Täuschung nicht einmal vorstellen können. In der Bewusstmachung dieser Gefahr liegt die Aufgabe der Täuschergotthypothese und in-sofern entspricht diese nicht der Supposition, wir seien Hirne im Tank, sondern der Supposition, es sei – entgegen unseren gerechtfertigsten Annahmen – eben trotzdem *möglich*, dass wir Hirne im Tank sind.

4.3.4 The Evil Demon

discussion of the basing relation

4.3.5 The Cogito

Chapter 5

Answering the Skeptic

5.1   Disbelieving the sceptics without proving them wrong

Abstract

It is true of many truths that I do not believe them. It is equally true, however, that I cannot rationally assert of any such truth both that it is true and that I do not believe it. To explain why this is so, I will distinguish absence of belief from disbelief and argue that an assertion of "*p*, but I do not believe that *p*" is paradoxical because it is indefensible, i.e. for reasons internal to it unable to convince. A closer examination of the irrationality involved will show that such is the skeptic's predicament, trying to convince us to bracket knowledge claims we have good grounds to take ourselves to be entitled to. Even if the sceptic cannot be proven wrong, his challenge still demands an answer, if not a treatment. In this paper, I argue that the cure lies in epidemiology rather than epistemology: instead of attacking the sceptic head-long, I commend guerilla tactics, vaccinating our fellow non-sceptics against the sceptical virus. I will not argue that the sceptic is wrong, necessarily wrong or that he cannot be believed, but that he cannot convince. Scepticism requires a leap of faith: something we may justifiably refrain from even on the sceptic's own standards.

5.1.1 Moore's paradox and pragmatic indefensibility

The puzzle of 'Moore's paradox', as I understand it, is to explain why the following sentence, while true for many values of p, is inappropriate to assert (Moore 1912: 78):

(5.1) p, but I do not believe that p

I will discuss two approaches to the puzzle which identify conversational maxims they take to be flouted by (5.1), criticise them and then present an alternative solution. The problem these diagnoses address is how to derive a contradiction from the claim that (5.1) is asserted, believed or entertained, i.e. to identify the pragmatic maxims and theoretical assumptions violated by such an assertion, belief, or act of entertaining.

What I will call the *belief approach*, exemplified by Hintikka (1962: 67), takes (5.1) to violate the maxim that one should believe what one asserts and then argues that (5.1) is unbelievable. To do this, it is argued that to believe a conjunction is to believe its conjuncts and that believing that one does not believe that p entails that one does not believe that p, i.e. that it is not possible to err about what one fails to believe (Hintikka 1962: 24).[1] If I believe (5.1), then, I believe that p and at the same time I believe that I do not believe that p, hence, by the 'veridicality of negative introspection', I do not in fact believe that p. In asserting (5.1), then, I assert something I cannot rationally believe and therefore act in a less than completely rational way.

What I will call the *knowledge approach* takes (5.1) to violate the 'knowledge rule' of assertion (Williamson 2000: 243), that one should only assert what one knows[2] and then argues that (5.1) is unknowable, on the assumption that I know a conjunction only if I know its conjuncts. If I know (5.1), then, I know that p and I know that I do not believe that p. Because knowing implies believing and by the factivity of knowledge, I hence both believe and do not believe that p, which is impossible (Williamson 2000: 254).

Both approaches have problems. After having sketched what I take to be their major ones, I will argue that the paradoxicality of (5.1) lies in its *indefensibility* and hence give a dialectical rather than purely pragmatic analysis of Moore's paradox. (5.1) is indefensible, I will argue, because any attempt to defend the claim made by it against an opponent contradicting p undermines whatever justification one might have had to assert "I do not believe that p" in the first place.[3]

The belief approach is committed to the claim that it is improper to falsely assert that one lacks a belief. But it seems clearly possible to believe that one lacks beliefs one in fact has. Just think of superstitions, prejudices and racial stereotypes, not to speak of unconscious beliefs. And such beliefs may be expressed with assertoric force. Even if we grant that we know of what we believe that we believe it,[4] this gives us only contradicting beliefs about one's own beliefs on the part of someone who asserts (5.1), which certainly is not enough to make the assertion improper.[5]

Another, related problem is that the belief approach is unable to explain − by itself −[6] why we do not only find (5.1) improper to assert, but (5.2) as well (cf. Williamson 2000: 253):

(5.2) p, but I do not know that p

---

1. This principle is a theorem of **KD4** and hence of **S4**.

2. The knowledge rule has also been defended by Récanati (1987: 183).

3. Hintikka (1962: 72) argues for the unbelievability of (5.1) from its doxastic indefensibility. I accept the premiss, but reject the analysis of belief that makes the transition valid. By calling my account 'dialectical', I do not mean to imply that only public utterances of (5.1) are paradoxical − any use of (5.1) as a premise in practical or theoretical reasoning counts as dialectical in the sense intended.

4. This principle has been defended e.g. by Halpern (1996: 487). That it carries substantial rationality presuppositions is shown by a quote of the present US president: "I know what I believe. I will continue to articulate what I believe and what I believe −I believe what I believe is right." (Rome, July 22, 2001)

5. If someone believes (5.1) he knows, on this account, that he believes that p and hence believes that he believes that p, while he also believes the second conjunct, i.e. that he does not believe that p. So he has contradictory belief which, however, is not enough to make him irrational.

6. This qualification is important: Hintikka (1962: 78-79) explains the unknowability of (5.2) by the **S4**-principle for knowledge. But this means giving a *different* account, basically Williamson's.

If for an assertion of (5.2) to be appropriate, I merely have to believe it, then it only follows that I have contradictory beliefs about my knowledge if believing that *p* entails believing that one knows that *p*. This principle, however, is itself of dubious standing as the paradox of the preface shows.[7]

The paradoxicality of (5.2) is elegantly explained by the knowledge approach.[8] A proper assertion of (5.2) would entail that I know that *p* and know that I do not know that *p* – which, knowledge being factive, is impossible. The knowledge approach has problems, however, with a different aspect of (5.1): not only is its assertion improper but it is also irrational to believe it. The standard reply to this worry is that knowledge sets the standard of appropriateness not only for assertion but also for belief (Williamson 2000: 47, 256). But if believing that *p* is to treat *p* as if one knew that *p*, then surely there is *some* mental state involving *p* that is not governed by the norms of knowledge. Let us call this mental state "entertaining".[9] Consider now the following assertion:

(5.3)   *p*, but I do not entertain (the thought) that *p*

(5.3) seems even more improper than (5.1) or (5.2) and the knowledge approach seems unable to explain this, for knowing that *p* does not imply entertaining that *p*.[10] Neither is it true in general that asserting implies entertaining. Assertions may be deferred and hence be made at different times than when the asserting sentences are uttered.[11] In such cases, the knowledge rule has to apply at the time of the assertion, of which I may know at the time of utterance that I do not then have any occurrent thoughts at all (because I am sleeping or dead, for example). The case is even clearer when (5.3) is merely entertained, mere entertainment not being governed by the knowledge rule at all. Such cases strike us, I assume, as equally puzzling than assertions of (5.1) and the knowledge approach does not seem to have the resources to explain our sense of puzzlement. If we combine it with the 'belief rule' (that one should believe what one asserts, believes or entertains), however, it forces us to rely on the dubious premise of the belief approach that negative introspection is infallible.

I would like to develop an account of the paradoxicality of (5.1) that presupposes only that one should believe what one asserts and does not rely on the dubious principle that one cannot have false beliefs about what beliefs one lacks. To do so, we need a distinction between absence of belief and disbelief.

5.1.2   Belief and disbelief

For any proposition *p*, to believe that *p* is to take the actual world to be a possible world where *p* is true.[12] Some possible worlds are, given what we take ourselves to know, ways the actual world might be; they correspond to informational states in which some of (what we take to be) our ignorance is remedied, ways of filling up gaps in (what we take to be) our knowledge.

What we take ourselves to know, however, is what we believe: Williamson, I think, is right that belief is governed, in this sense, by the norm of knowledge. This is why I call such informational states *doxastic alternatives* (for some agent at a time). Not all doxastic alternatives are possible worlds: sometimes we think we know propositions which

---

7. The paradox of the preface shows that it may be rational to think that at least one of one's own knowledge claims is false, as it may be rational to excuse oneself, in the preface, for one's errors in the book, thereby implying that there are some. One believes all the claims one makes in the book, but the remark in the preface shows that one does not think one knows them all. Dretske (1971: 21) even denies that *knowing* that *p* entails believing that one knows that *p*: "We naturally expect of one who knows that *P* that he believes that he knows, just as we expect of someone who is riding a bicycle that he believe he is riding one, but in neither case is the belief a *necessary* accompaniment."

8. This is even acknowledged by Hintikka who, in order to explain the inappropriateness of (5.2), appeals to something like Williamson's knowledge rule of assertion, though he considerably weakens it: "When somebody makes a statement […] we are normally led to expect that he can conceivably know that what he is saying is true or that he is at least not depriving himself of this possibility by the very form of words he is using." (Hintikka 1962: 78). Hintikka has, however, a much more internalist conception of what it takes to be a norm of assertion than Williamson. Cf.: "It cannot very well be a presupposition of the notion of honest assertion that the speaker knows what he says; one may be honestly mistaken." (Hintikka 1962: 97)

9. Entertaining is different from opining in that only the latter is supposed to capture whatever else than knowledge is needed for a disjunctive definition of belief (Williamson 2000: 46).

10. Instead, it gets the hierarchy wrong: (5.2) is most directly classified as inappropriate, (5.1) indirectly via the route from knowledge to belief and (5.3) via a dubious connection between knowledge and entertaining. It seems to me, though, that (5.3) is the most paradoxical of all three sentences considered.

11. Such cases have been widely studied in the literature on indexicals: think of post-it notes on your office door saying "I am not here now", messages on answer-machines and so on.

12. I use "proposition" to stand for whatever it is that may be known, believed, true and expressed by sentences. For the purposes of this paper, I individuate them up to logical equivalence.

in fact are impossible.[13] Nor do I take the valuation to be total: doxastic alternatives typically do not decide all issues. On the model sketched here, we may be ignorant of necessary and of a priori truths (and even know that we are).[14]

To believe p is for p to be true in all one's doxastic alternatives; it is to answer "yes" to both the question whether p and the question whether one believes that p. Not every failure to believe, however, is a case of disbelief: disbelief is an epistemic attitude, albeit a negative one, and there may be many propositions to which I have no epistemic attitudes at all. To disbelieve that p is to consider it possible that a world which verifies $\neg p$ is actual, to have a doxastic alternative in which $\neg p$ is true; it is to answer a firm and non-reflective "no" or "of course not" to the question whether p (as opposed to the deliberation of whether you should believe it or not and making up your mind on the matter). Disbelief that p, then, is stronger than mere lack of belief. It is weaker, however, than belief that $\neg p$.[15] To believe that $\neg p$, all of one's doxastic alternatives have to make p false (i.e. $\neg p$ true); to disbelieve that p, one doxastic alternative in which $\neg p$ is true suffices.

To see the difference between the belief that $\neg p$ and the disbelief that p, consider the following: Whenever someone asks me whether I believe that p, there are two ways of giving a negative answer: merely stating my lack of belief, as in "no, I never thought about that", and expressing a negative epistemic attitude to the proposition that p, as in "no, of course not, that would be idiotic". This negative attitude, however, may itself be of two kinds, distinguished by different commitments. Suppose now someone asks me a second question, namely whether p. Again, there are two ways of answering in the negative: I may, first, express my belief that $\neg p$, but I may also just reject the proposal that p. It is this second way of answering "no" to "p?" that is equivalent to the negative epistemic attitude of disbelief with respect to p. Doxastic alternatives provide a neat way of bringing out these distinctions: they allow us to distinguish between someone who incurs a commitment to $\neg p$ and someone who refrains from (and not merely lacks) a commitment to p.

Though disbelief is distinguished from lack of belief by its being an occurrent epistemic attitude, it may also be construed dispositionally: Disbelieving that p is to have the disposition to refrain from acting on p, which is not equivalent to and weaker than the disposition to act on the falsity of p – but stronger as lacking any disposition with respect to p altogether. Suppose the waiter brings me a dish x I did not order. It seems true to say that at the time of ordering my dish, I did not want to eat x (today). This absence of a pro-attitude may be of two sorts, however: it may be a mere absence, or an absence due to the presence of a contra-attitude. In the first, but not the second case, I can cleave to my former preferences by taking the dish.

The notion of disbelief is needed to account for agnosticism: the typical agnostic, e.g. with respect to religious matters, is not just open-minded (lacking both the belief that p and the belief that $\neg p$), but he is not committed to dialetheism (believing both that $\neg p$ and that $\neg(\neg p)$). Agnosticism with respect to p, instead, is an epistemic attitude characterised by disbelief that p and disbelief that $\neg p$ – this is why agnosticists typically have arguments for their position that have to be overthrown to convince them of either p or $\neg p$: they have doxastic alternatives to be ruled out by argument.

Disbelief, in my view, is what underlies truly 'rejective' negation. This is not, however, the use of this word that became prevalent in recent discussions (cf. Rumfitt (2000) and Humberstone (2000)). Like Frege[16] and Dummett,[17] Rumfitt treats rejection and holding-false as equivalent:

---

13. Equivalently, on the model sketched here: we believe contradictory propositions (without, in normal cases, realising that we do so).

14. Neither do I assume that any agent in any world either believes or disbelieves a given proposition: there might be worlds, e.g. those in which a does not exist or those in which a lacks the relevant concepts, where a neither believes nor disbelieves that p. Neither do I see a reason why a should not be able to consider such worlds to be doxastic alternatives.

15. Some terminological regimentation is involved here. For a different use of the word, cf. e.g. Price (1983: 171) according to whom "to believe that S is to be disposed to act as if S, whenever one thinks it makes a difference whether S" and "to disbelieve that S is to be disposed to act on the assumption that not-S – that is to ignore the possibility that S – whenever one thinks it makes a difference to the outcome of one's actions whether S".

16. Cf. in the early *Logik*: "In asking a yes-or-no question we are wavering between opposite sentences. […] This opposition or conflict is to be understood in such a way that we *ipso facto* reject one limb as false when we accept the other as true, and conversely. The rejection of the one and the acceptance of the other are one and the same." (Frege 1879–1891)

17. Cf.: "…the answer 'No' is precisely tantamount to, and is best analysed as, an assertion of the negation of the sentence uttered in interrogative form" (Dummett 1973: 425).

> On the bilateral conception [the one Rumfitt advocates], a speaker is entitled to affirm a sentence's negation precisely when he is entitled to reject the sentence itself. (Rumfitt 2000: 818, cf. also 814)[18]

"Rejecting the sentence" here means answering negatively the question whether the sentence is true, in the sense in which this expresses belief in its negation. The issue is somehow complicated by the fact that for Dummett, Rumfitt's main target, it is correct to 'reject' (deny) A iff it is correct to assert $\neg A$ iff it is correct to assert that A is not assertible. It is only on the premiss of the second equivalence that Rumfitt denies the right-to-left direction of the first: he disagrees with Dummett on whether if it is incorrect to assert A, then it is correct to reject A.[19] Because Rumfitt concedes that if it is correct to reject A, then it is correct to assert that $\neg A$, and that the latter implies that A is false, he has to claim that it may be incorrect to assert A even if it is not the case that A is false; i.e., accepting bivalence, that falsity is not sufficient for correct 'rejection' (denial) and hence that the intuitionist is wrong in construing an assertion of $\neg A$ as warranted iff the assertion is warranted that nobody will ever be in a position to make a warranted affirmation of A.[20]

I want to claim, however, that it may be correct to reject that p even if it is not correct to assert that $\neg p$, precisely because 'rejection', as opposed to denial, is best construed as a manifestation of disbelief. This does not mean that I deny the shared premiss of both Dummett and Rumfitt that there are no conditional assertions, i.e. that assertions are either correct or incorrect. It is just to claim that a rejection of p is not equivalent to, and hence not to be construed as, an assertion of $\neg p$.[21] Its 'assertoric' force may be weaker.

A doxastic alternative for a at t, then, is a way the world might be a cannot exclude at t to be a way the world actually is.[22] If a believes that p, all his doxastic alternatives make p true; if he disbelieves p, one of his doxastic alternatives makes p false, i.e. makes $\neg p$ true. We now have the following dual pair of epistemic notions, where "**Bel** p" abbreviates "a believes that p", "**Dis** p" stands for disbelief and "$\mathscr{DA}$" for the relation of doxastic alternativeness with respect to a and t:

(5.4) $w \models \textbf{Bel}\, p \quad :\Leftrightarrow \quad \forall v(w \mathscr{DA} v \to v \models p)$

(5.5) $w \models \textbf{Dis}\, p \quad :\Leftrightarrow \quad \exists v(w \mathscr{DA} v \land v \models \neg p)$

As argued above, I understand disbelief in p as an epistemic attitude which is, in normal cases, intermediate in strength between mere lack of belief and belief in the negation. If one even believes both p and $\neg p$, all of one's doxastic alternatives are contradictory. Disbelieving that p entails not believing that p if the relevant doxastic alternative is not contradictory. For the entailment from failing to believe that p to disbelieving that p, however, something much more stronger is needed, namely that the relevant doxastic alternative makes p false if it does not

---

18. Consequently, Rumfitt's system has it that $+(\neg A) \dashv\vdash +A$ (where $\ulcorner +A \urcorner$ represents the assertion of A and $\ulcorner -A \urcorner$ its rejection, or "denial", as Dummett (2002: 290) – and subsequently (Rumfitt 2002) – prefer to call it).

19. Cf.: "What he [Rumfitt] disputes, and I [Dummett] have maintained, is that the negation of a sentence may be asserted – or the sentence denied – just in case we have good reason to judge that it would be objectively incorrect to assert that sentence." (Dummett 2002: 291) The example Rumfitt (2000: 818) gave of such a sentence was "Elizabeth I was bald when she died" in the situation where all documentary evidence pertinent to her person has been (and is known to have been) collected and destroyed. Dummett (2002: 292) denies that this is a counterexample: "We are surely in no position to deny that *there ever have been* grounds for asserting the sentence [...], and are therefore not [...] in a position to deny that sentence." But just suppose we were (being told so by God, say): plainly, this would not, by itself, justify an anti-realist about the past in asserting the negation (and this was Rumfitt's original point).

20. It is precisely *because* he construes denying a sentence as asserting its negation that Rumfitt claims that "[c]onfronted with the question $\ulcorner$ Is it the case that A?$\urcorner$, an adherent of the bilateral conception will insist upon distinguishing between the responses "There are no grounds for answering affirmatively" and "No"" (Rumfitt 2000: 798), and that there are fields of inquiry where "there are now, never have been and never will be grounds for answering affirmatively" is weaker than "no" and which therefore sustain a "genuinely bilateral conception of sense" and favour classical over intuitionistic logic (Rumfitt 2000: 817). Only if the falsity of a sentence is then identified with its being correct to deny it, we get the paradoxical result that classical logic is correct in areas of discourse in which bivalence fails (cf. Dummett 2002: 293).

21. This, I think, is needed to meet the challenge set up by Humberstone (2000: 368), namely "to show how the claim for the conceptual priority of rejection over negation is any more plausible than the corresponding claim for the conceptual priority of alterjection [a speech-act whose propositional objects are two disjuncts and which expresses a 'disjunctive attitude' with respect to two propositions] over disjunction".

22. I assume that the only way to exclude some world from the realm of informational states in which some of our ignorance is remedied is by self-ascribing some putative pieces of knowledge that fail to be true in that world. Some care is needed here, however: we may (rightly and justifiedly) take ourselves to be ignorant about many things and thus there had better be some doxastic alternatives where our first-level ignorance is remedied and we therefore lack the second-order bit of knowledge. This may be assured in different ways, e.g. either by restricting the claim to first-order knowledge or to time-index the knowledge claim (such that it is not *that* bit of knowledge we later lack, for it only concerned what we knew at the other world and the earlier time). I thank Timothy Williamson for having me made aware of this.

make it true. Failing to disbelieve $\neg p$ is to have only doxastic alternatives which do not make p true, i.e. being open-minded about the truth of p. Failing to disbelieve that $\neg p$ is the sharpest contrast to believing that p and means that one has no doxastic alternatives, i.e. that there is no way in which one's belief state could be made true compatible with what one takes oneself to know: it is to answer both "yes" to the question whether p and fail to answer "no" to the question whether $\neg p$. Believing and disbelieving that p (answering both "yes" and "no" to the question whether p), on the other hand, means that one has a doxastic alternative, though a contradictory one. This shows that doxastic alternativeness is a highly intensional notion, expressing not more than what *a takes to be* the relevant scenarios to consider. Even if a's belief set is inconsistent and he thus has doxastic alternatives which both verify and falsify some proposition, we cannot attribute him any belief whatsoever: to do this, he must positively fail to grasp the force of the negation.[23]

5.1.3 Pragmatic indefensibility

How does this machinery help us with (5.1)? If we construed doxastic alternativeness as a transitive relation, it would be straightforward to show that Moore's paradox is unbelievable.[24] This was essentially the solution given by Hintikka's **S4**-logic for knowledge and belief, construed in terms of defensibility of belief sets and satisfying the following principle of positive introspection:

(4) $\quad \vdash \textbf{Bel}\, p \rightarrow \textbf{Bel}\,\textbf{Bel}\, p$

(4) seems implausible for reasons already mentioned: superstitions, prejudices, unconscious, externalistic and essentially indexical beliefs. Another reason against it is the following: if it were adopted, it would be quite difficult to resist the temptation of giving a similar account of knowledge. Justification, after all, is transitive and it seems difficult to see what would block (4) for knowledge other than lack of the relevant belief. And I take it that (4) is extremely implausible for knowledge.[25]

A diagnosis of (5.1) in terms of the transitivity of doxastic alternativeness has a further drawback. What makes the indefensibility of (5.1) paradoxical is not only that we know that, for any p, it might be true and for most of the true p's of our language it actually *is* true. It also is paradoxical because there does not seem anything wrong with ascribing lack of belief in a true proposition *to someone else* or to oneself *with respect to some other time than the present*.

---

23. If the agent both believes p and believes $\neg p$, *all* his doxastic alternatives will be contradictory, if he has any. Only if he does *not* have any, i.e. if he answers not only "yes" both to p? and $\neg p$? but also fails to answer "no" to the second question, thereby showing that he failed to understand that he is contradicting himself, may we safely take him to be irrational (think of someone who answers not only both "yes" and "no" to the question whether God is immortal but also refrains from answering "no" to the question whether he is mortal – we may well wonder whether he understood what "immortal" means). Though believing p to be false (believing that $\neg p$) is stronger than failing to disbelieve $\neg p$, belief in p and failing to disbelieve $\neg p$ is a stronger form of irrationality than both believing p and believing $\neg p$. The connection between belief and disbelief in the negation is internal and constitutive, while both belief in p and in its negation may be independently justified. Disbelieving both p and $\neg p$, in contrast, is entirely unproblematic and just requires doxastic alternatives verifying p and (possibly different) doxastic alternatives falsifying p.

24. Assume there is a possible world w where a believes that p and that a does not believe that p: $w \models \textbf{Bel}\,(p \wedge \neg\textbf{Bel}\, p$. Assume there is a doxastic alternative v for a in w ($w\mathscr{D}\mathscr{A}v$). Then $v \models p \wedge \neg\textbf{Bel}\, p$. Because of the second conjunct, there is an doxastic alternative u for a in v such that $u \not\models p$. By transitivity, u is also a doxastic alternative for a in w. Hence $u \models p$, which is impossible. So a has no doxastic alternatives in w, that is, may safely be taken to be irrational.

25. The strongest argument against (4) in the case of knowledge, I think, is the simple fact that knowledge may supervene on factors unknown to the agent: Suppose that whether or not I have knowledge of the truth that p depends on the obtaining of some external factors r (i.e. a reliable connection between my belief and the fact that p, the trustworthiness of my conversational partner, the actual absence of other possibly misleading evidence, subjunctive conditionals connecting my believing and not-believing that p with p and $\neg p$ respectively, or what have you) such that, if r were not the case, I would not know that p. It is enough, then, for my knowing that p that r obtains – I do not have to know that it obtains. In order to know that I know that p, however, I have to rule out that $\neg r$, which, in many cases at least, demands for a different type of cognitive effort than establishing whether p. An especially vivid case is where I even lack the concepts or linguistic resources to express r: it seems not clear, in this case, how I might be able to rule out that $\neg r$, even while I do not have to articulate the background conditions on which may ordinary, first-level knowledge claims depend. Timothy (Williamson 2000: 13, 95) frames the point in terms of luminosity, where a mental state is luminous iff whenever it obtains, one is in a position to know that it obtains. Knowledge is not luminous and hence (4) fails, according to (Williamson 2000: 107), because it can be gained or loosed gradually, without the subject being able to draw a clear-cut line along the slippery slope: that is why knowledge requires "margins for error", inside which (4) fails. Transitivity is one of the respects in which information, "travelling at the speed of logic", most perspicuously differs from "genuine knowledge [which] only travels at the speed of cognition and inference." Barwise (1988: 204). Transitivity of information flow has been taken as axiomatic by Dretske (1981: 57): "I take this [the Xerox principle, i.e. transitivity of information flow] to be a regulative principle, something inherent in and essential to the ordinary idea of information, something that any *theory* of information should preserve. For if one can learn from A that B, and one can learn from B that C, then one should be *able* to learn from A that C (whether one *does* learn it or not is another matter)."

Using transitivity to show that Moore's paradox is unbelievable *tout court* either misses this feature or bars us from many interesting applications:[26] the doxastic alternatives to consider typically lie in our future; the doxastic agents we have beliefs about are typically not ourselves, though we may become them if we upgrade things we believe to things of which we believe that we know them.[27]

Jaakko Hintikka argued for (4) as an axiom of knowledge on the basis of his peculiar semantics given in terms of defensibility (consistency) of belief sets:

> That q is the case can be compatible with everything a certain person – let us assume that he is referred to by a – knows only if it cannot be used as an argument to overthrow any true statement of the form "a knows that p." Now this statement can be criticised in two ways. One may either try to show that p is not in fact true or else try to show that the person referred to by a is not in a position or condition to know that it is true. In order to be compatible with everything he knows, q therefore has to be compatible not only with every p which is known to him but also with the truth of all the true statements of the form "a knows that p." (Hintikka 1962: 18)[28]

The idea, in a nutshell, is that I only know something if its truth is not excluded by what I know nor by my knowing what I know. This is based on the observation that explicit self-ascription of knowledge is a particularly strong form of endorsing a claim:

> I am not in a position to say "I know" unless my grounds for saying so are such that they give me the right to disregard any further evidence or information. (Hintikka 1962: 20)

This argument from our dialectical entitlement to knowledge claims[29] is inconclusive, however. It may be defended (at least by Hintikka-style arguments) only in contexts where the "absurdity" correlated to an indefensible belief set is understood in performatory terms.[30] This performatory aspect of knowledge claims, however, is cancellable and applies only to first-person utterances. Whenever we are interested in modelling the epistemic behaviour of (real or idealised) agents, such considerations are inapplicable. This is why (4) is not an axiom we should adopt in our logic of knowledge or belief.

Hintikka's considerations mooted above, however, point to an important feature of self-ascriptions of belief or knowledge. Whenever we say of ourselves that we believe that p, we incur a commitment to the truth that p – we claim that the real actual world (not only the one we take to be the actual world) is among our doxastic alternatives. Whenever we learn further truths and acquire true beliefs, we narrow down the range of alternatives, hopefully to an ideal limit where the actual world would be our only last alternative left and we would believe some proposition iff it is true.[31] A self-ascription of knowledge or belief is a claim to the effect that we are prepared to use some

---

26. Hintikka only considers statements made on one and the same occasion and presupposes that the person referred to by "a" in the index of the knowledge operator knows that he is being referred to by "a" (Hintikka 1962: 106) – a piece of knowledge that cannot itself be modelled within the system.

27. This is not the only way a doxastic alternative may become actual, of course: any change in what I take myself to know defines a new doxastic alternative.

28. He had an ancillary argument for transitivity, based on the observation that "knowing" and "knowing that one knows" are often used interchangeably in ordinary language – or, at least, that their "basic meanings" as captured by an explanatory model are the same (cf. Hintikka 1968: 8).

29. Hintikka sometimes also puts the point in terms of justification: "In the primary sense of *know*, if one knows one *ipso facto* knows that one knows. For exactly the same circumstances would justify one's saying "I know that I know" as would justify one's saying "I know" *simpliciter*." (Hintikka 1962: 28, cf. also 111)

30. Hintikka is clear on this point: "The absurdity of doxastically indefensible sentences is of *performatory character*; it is due to doing something rather than to the means (to the sentence) which is employed for the purpose." (Hintikka 1962: 77) He calls transitivity also "the quasi-performatory aspect of the verb *know*" (Hintikka 1962: 55) and repeatedly stressed this point in subsequent discussions: "[…] for someone to know that p his evidence […] has […] not only to be good but as good as it […] can be. It has to be such that further inquiry loses its point (in fact, although it is logically possible that such an inquiry might make a difference). The concept of knowledge is in this sense a 'discussion-stopper'. It stops the further questions that otherwise could have been raised without contradicting the speaker." (Hintikka 1968: 13)

31. Again, this has to be restricted to non-epistemic propositions. Even then, however, it may be questioned whether ideal believers at the limit of inquiry should believe that they were ideal believers at the limit of inquiry. These matters, interesting as they are, are orthogonal to my purposes here: all I need is that it is irrational to exclude the real actual world from one's doxastic alternatives, whatever is the best way of spelling out the epistemic norms at work here.

proposition as a premise in this process of narrowing down the realm of what is left open by what we take ourselves to know. A crucial feature of beliefs we claim to have, then, is that they must entertainable in worlds which are their only doxastic alternative. Any proposition that cannot be believed in such a world immediately disqualifies from the dialectic role conferred to it by an act of self-ascribing belief in it.

What lies at the bottom of indefensibility of (5.1) is not the transitivity of doxastic alternativeness but the commonly made presupposition that what one says might be true even if the real actual world were one's only doxastic alternative. If I assert that p, I thereby commit myself to the claim that p might be true even if everything I believe would count as knowledge – and this requires that the actual world is among my doxastic alternatives.[32] If you assert p, you must consider it possible that you would believe p even if you had a maximally specific belief set, i.e. if you would believe all the (non-epistemic) truths (or, equivalently, disbelieve the infinite conjunction of all the (non-epistemic) falsehoods).

This maxim allows us to bring out the indefensibility of (5.1). If the world in which we believe (5.1) were our only doxastic alternative, belief in (5.1) would make that world inaccessible to us. A belief in (5.1) makes the world in which it is held either doxastically inaccessible to itself or contradictory. Therefore, (5.1) cannot be rationally believed in 'reflexive' worlds, i.e. worlds w such that $w \mathscr{D} \mathscr{A} w$.[33] How dramatic is this?

To answer this question, we have to distinguish between two sorts of epistemic possibility, corresponding to an externalist and an internalist reading of the phrase "the actual world" in "a way the actual world could have been (given what we know about it)". According to its externalist interpretation, anything which is compatible with our knowledge and hence might be, for all we know, a world we happen to inhabit qualifies as an epistemic possibility. Anything we cannot rule on the basis of what we know about *this* world is possible in this first sense, even if it would preclude this world's being a world we inhabit. Because there is no knowledge we are entitled to rely on when arguing with (that is, against) the skeptic, any not obviously self-contradictory scenario will qualify as a way the actual world could have been, taking the latter in its referential sense. An internalist reading, however, demands more: to be epistemically possible in this second way, a possible world not only has to be one of which we do not know that it is not ours, but more, namely one which we could *take to be* our actual world. For a skeptical scenario to depict a possible world in this second sense, then, is for it to be such that it could, in principle, be taken to describe our actual world. For this it has to be compatible not just with what we know, but with what we take ourselves to know. This distinction, between what we may call *merely possible* and *possibly actual* worlds, will turn out crucial in the following.

I tried to capture the intended internalist reading by relativising doxastic alternatives to what an epistemic agent *takes* himself to know and by requiring that a rational agent should have self-accessible doxastic alternatives. Claims which are true only in possible worlds which are not possibly actual are plausibly taken to be *indefensible* because they cannot convince anyone who does not believe them already: it is only by a leap of faith that you can come to believe them to be actual. For to take them to be possibly actual *is* to take them to be doxastic states one rationally actually might be – states which are compatible with what is taken to be true in them.

Taking a doxastic alternative to be possibly actual as opposed to just merely possible, then, means taking it to be doxastically accessible to itself. So no one can take a doxastic alternative in which (5.1) is true to be a possible way *his* actual world, the world of the believer, might be. This is what is paradoxical about the Moore-paradoxical sentence (5.1): that it cannot, on pain of contradiction, be believed in worlds which are doxastic alternatives of themselves.[34]

It is instructive to compare (5.1) to the following more explicit version:

(5.6) I believe that p, but I do not disbelieve that $\neg p$.

---

32. Something along this line was Hintikka's original justification for the transitivity of his belief relation: "If something is compatible with everything you believe, then it must be possible for this something to turn out to be the case together with everything you believe without making it necessary for you to give up any of your beliefs. If your beliefs are to be consistent, it must also be possible for all your beliefs to turn out to be true without forcing you to give up any of them." (Hintikka 1962: 24)

33. Suppose $w \models \mathbf{Bel}\,(p \wedge \neg \mathbf{Bel}\,p)$ and $w \mathscr{D} \mathscr{A} w$. Because of $w \models \mathbf{Bel}\,\neg \mathbf{Bel}\,p$, $w \models \neg \mathbf{Bel}\,p$. But the first conjunct becomes $w \models \mathbf{Bel}\,p$, so w is contradictory.

34. This diagnosis equally applies to (5.2) and (5.3). In the latter case, this is due to the fact that at the ideal end of inquiry you would entertain all and only the true of the non-epistemic propositions.

As was said before, asserting (5.6) amounts to the avowal that one's belief set is inconsistent. (5.6) can be believed only in worlds which have only worlds as doxastic alternatives which themselves have no doxastic alternatives.[35] (5.6) is a stronger assertion than (5.1), for not every sentence which I non-contradictorily fail to believe I automatically disbelieve. If I non-contradictorily disbelieve something, however, I fail to believe it.

Another case is the following:

(5.7) p, but I do not disbelieve that $\neg p$.

(5.7) can only be believed in worlds which only have doxastic alternatives that are doxastically inaccessible to themselves.[36] The pragmatic maxim violated by an utterance of (5.7) is not, as in the case of Moore's paradox, that one should believe what one says but something weaker, i.e. that one should disbelieve what one believes to be false. It is thus unbelievable or unassertable in a stronger sense than the Moore-paradoxical sentence (5.1). The believer in (5.7) cannot take the world he takes to be actual (and therefore accessible to itself) to be one of its doxastic alternatives, not even, as in the case of (5.1), at the price of acknowledging that it is contradictory. He cannot take any world to be its doxastic alternative without foresaking that it may be possibly actual (for if it *were* actual, he could not take it to be actual, for he would then have a doxastic alternative which is accessible to itself). I think it is this feature of (5.7) that explains our being bemazed by an assertion of (5.1), for we naturally interpret an assertion of lack of belief as an assertion of disbelief.

5.1.4   Disbelieving the skeptic without proving him wrong

I think that this interpretation of (5.7) helps us tackle what I still take to be the main problem of epistemology: not whether the skeptic is right but why he is wrong. Skepticism seems to have found its niche in philosophical multiculturalism. While still unfashionable, it has become tolerated, or rather ignored. I think skeptics earn better than that. Even given that they cannot be proved wrong, their challenge still demands an answer, or rather a treatment. I will argue that the cure to skepticism lies in epidemiology rather than epistemology: instead of attacking the skeptic head-long, I commend vaccinating our fellow non-skeptics against the virus. The way to go is not to argue that the skeptic is wrong, necessarily wrong or that he cannot be believed, but to show that he cannot *convince*. Skepticism requires a leap of faith: something we may justifiably refrain from even on the skeptic's own standards – or so I want to argue.

There are very few skeptics. The importance of skepticism to epistemology is not due to its popularity. Instead, epistemologists cherish skepticism because it provides a convenient setting to many important epistemological questions and a challenge to some of their most popular answers. The issue of skepticism, then, is not so much the question whether the skeptic is right, but *why he is wrong*. So let us assume, at least for the purposes of this paper, that *we* know that the skeptic is wrong and set us the task to explain and defend our entitlement to this knowledge claim.

Skeptics are not very good at mounting positive claims. Instead, they ask questions. I take the skeptic of concern to us to be someone who believes that no one ever knows anything. As we assume him to be rational, he will not qualify this belief of his as knowledge. The skeptic thus thinks that all our beliefs that we know something are false. The question he asks us is the following: "How do you know that you know something?" By taking his question to be rhetorical, he challenges our entitlement to any knowledge claims (claims of the form "I know that p") whatsoever. To meet the skeptic's challenge, we have to develop an account of knowledge that gives us the resources to defend against the skeptic the claim that we know something.

Just by knowing something, we only show *that* the skeptic is wrong. In epistemology, however, this will not do: as everywhere else in philosophy, we have to give reasons for taking our knowledge claims to be more than socially useful fictions. So it will not do to just know something, we have to know that we know it. Even this claim, however, demands for justification. To meet the skeptical challenge in full generality, then, we would have to

---

35. Suppose that $w \models \mathbf{Bel}(\mathbf{Bel}\,p \wedge \neg\mathbf{Dis}\,\neg p)$ and there is a v such that $w\mathscr{DA}v$. Then $v \models \mathbf{Bel}\,p$ and $v \models \neg\mathbf{Dis}\,\neg p$. If $v\mathscr{DA}u$, then $u \models p$ and $u \not\models p$, which is impossible.

36. Suppose that $w \models \mathbf{Bel}(p \wedge \neg\mathbf{Dis}\,\neg p)$ and there is a v such that $w\mathscr{DA}v$. Then $v \models p$ and $v \models \neg\mathbf{Dis}\,\neg p$. If $v\mathscr{DA}v$, then $v \not\models p$ which is impossible.

defend something very much like the **S4** axiom, i.e. we would have to show that for everything we know, we know that we know it. To do this without begging the question against the skeptic (e.g. by assuming that we have a privileged access which gives us *knowledge* about all of our knowledge) seems to me very difficult. So let us simply assume that it cannot be done.[37] But even if would be done somewhen in the future, it seems worthwile to have a safe fallback position in the meantime.

The question of what to say to the skeptic lands us in a dilemma. The skeptic sketches a skeptical scenario, *p*, furnishes a description of what he takes to be a possible world in which we would not know what we claim to know.[38] He then challenges us to explain to him why we think we are justified in excluding that scenario, i.e. to exclude its possible actuality. Given that we steadily stay non-skeptics, we have to choose between two equally uncomfortable options: either we find a hidden contradiction in what the skeptic presents as possible (which has turned out very difficult, if not impossible in many cases) or else we simply declare our psychological inability to believe him and thereby end the discussion before it even got started.

A medium between these extremes, I think, is to take skepticism to be an epidemic: a contagious superstition, which, for fear of infection, we do better to isolate than to confront directly. Such a epidemiological rather than epistemological response requires some redefinition of what it means to meet the skeptical challenge: we do not have to convince those among us who are already infected with the skeptical virus; instead, we have to prevent the skeptical disease from propagating, i.e. to detain the skeptics among us of convincing others. In order to do this, we cannot stay with our own inability to believe the skeptics: we have to justify and to explain to others why we cannot get ourselves into believing them. Our inability to waive our knowledge claims has to be argued for – in a way that *shows* (displays, not demonstrates) that it is not idiosyncratic, not prejudiced and not just a symptom of our unwillingness to consider their arguments. To meet the skeptical challenge, therefore, I have to win a three-person game: I, the antiskeptic, have to convince you, the innocent bystander, that you should not believe the skeptic who is trying to convince you of the epistemic possibility of the scenario he sketches. I have to convince you that the reasons I have for not being a skeptic carry over to your case, *if you are not a skeptic already*. Taking skepticism seriously, then, is to take it seriously as a threat. It is not necessarily trying to refute it: it is enough to show that what the skeptic takes to be possible is not possible *for the two of us*.[39]

This strategy bears some resemblance to Harty Field's 'evaluativism', according to which "we *inevitably* believe that our own methods will be better than the dissimilar methods at leading to truths and avoiding errors" (Field 2000: 127). Evaluationism, however, is a stronger claim than the one argued for here: I do not conceive of the skeptic as proposing *different* epistemic policies but as trying to undermine our confidence in our own by asking the (supposedly unanswerable) question why we prefer it to some of the imaginable alternatives. To say that this is inevitable (according to our standards) is one, but not the only way, of saying that the only way of becoming a sceptic is by a leap of faith.

Some possible worlds are, given what we know, ways the actual world might be. We cannot exclude them when evaluating modal claims, for they do not contradict anything we know. Different amounts of knowledge we are allowed to take for granted give us a continuum of more and more inclusive contexts and different grades of

---

37. This perhaps prematurely pessimistic view has found some adherents in the recent discussion (cf., e.g., Stroud (1984: 20)) and seems at least not implausible on inductive grounds.

38. I do not take a skeptical hypothesis to be a scenario in which what we think to know is false, but one in which we do not know what we think to know (though it still may, unbeknownst to us, be true). I have three reasons for this: first, a skeptical hypothesis should describe the world as it is believed to be by a skeptic and a skeptic is not committed to the belief that the world *actually* is radically different than we take it to be, but only that our world might be so, given what we know. My second reason is historical: I believe that Cartesian skepticism (the kind of position that the protagonist of the First Meditation believes himself in) is not characterized by the actuality, but by the mere possibility of total delusion. Third, my notion is more general (for scenarios in which what we believe to know is false are scenarios in which we do not know it): I am thus making the skeptic stronger than others.

39. I am not claiming that this way of answering the skeptical challenge is preferable or even defensible under all circumstances. At the basis of the treatment of skepticism I am going to propose lie two brute empirical facts, one sociological and one ideological: first, there are very few skeptics (so doing without them is a viable option), second, the grimest of them are not among us, but in the madhouse. Doubting is a mental activity and can, like all mental activities, get out of control, i.e. it can acquire a causal role radically different from the one it has in most of us. A state of doubt which is insensitive enough to arguments applying at a stage posterior to the skeptical conversion qualifies as mad doubt in this sense. The skeptic who cannot, even for the sake of the argument, imagine himself being a non-skeptic, disconnects his state of doubting from its usual effect, i.e. reliance on commonly accepted procedures of finding out. I am not denying that this situation could (or perhaps even should) not be the inverse, we in the madhouse for treatment of mad confidence, they out there paying for us. Instead of trying to justify what we do anyway, I accept it as a brute fact and will try to show that and how it applies to the skeptics' case.

epistemic possibility. Contextualism is the view that our notion of knowledge has to be understood relative to some such set of presuppositions we are entitled to make. What we know depends, inter alia, on the context we inhabit at a particular time. What is good enough for knowledge in some situations, may well fall short of it in others. The skeptics' question forces us into a rather particular context. By raising the question whether we can exclude the possibility of a particular skeptic scenario *p*, the skeptic forestalls our ignoring it. He questions the validity of *all* knowledge claims we happen to make, so he cannot be disproved by relying on any of them. To take him seriously, we have to bracket anything we believe to know. Contextualism then is just a way of begging the question against the skeptic. This is not as bad as it sounds, given that the skeptic cannot be refuted. Still, we should be able to do better. And we can, by showing that the skeptical question *cannot be but* begged.

I want to argue that the best known scenarios skeptics and anti-skeptics have produced so far is, *for the innocent bystander to be convinced by the skeptic*, relevantly similar to the way Moore's paradox is *for us*: both of them are unable, by the very structure of their respective assertions, to produce conviction. Though they may be true and even believed to be true, the skeptic trying to convince us has to show more: he has to show that we are not epistemically required to disbelieve that they are actual. He has to show that the scenario he depicts is possibly actual, i.e. that we might take ourselves to be in it – and he has to do this without presupposing that we actually are in the unfortunate situation depicted by his scenario.

Disbelieving that *p*, on our account, is to have a doxastic alternative where ¬*p*. The skeptic who believes that noone ever knows something has to consider every possible world an doxastic alternative. This is clearly possible. What is not possible, however, is to convince somebody who already believes he knows something that a world like the one the skeptic takes to be actual is a doxastic alternative for him.

Assuming that the innocent bystander is not actually deluded (¬*p*), the skeptic has to show him that he does not disbelieve he is deluded. He has to show him that he even though he *is* not deluded, he may still refrain from answering "no" if asked whether he is deluded. He has to show him how to bracket his standing belief in non-delusion without assuming that he is actually deluded, i.e. that non-delusion (¬*p*) is compatible with failing to disbelieve that one is deluded (¬**Dis** *p*). So the skeptic has to convince us of (5.7), i.e. that we are not deluded, but at the same time do not disbelieve we are deluded. Only in this way he can overcome our cognitive resistance. But this task is hopeless.

The problem with (5.7) is not the inappropriateness of uttering it, but that belief in it cannot be taken to be true. The problem is that it can only be true *and* believed in worlds which are 'anti-symmetric', i.e. which are such that every doxastic alternative to them does not have them among its doxastic alternatives.[40]

Because convincing others of *p* requires not only that you take yourself to believe *p* but also that you convince them that they do not have to disbelieve things they already believe (though they might suspend their belief in some such things), such an argument would show that skeptics cannot convince others of skepticism. Instead of showing the skeptical scenario to be impossible, it shows that it cannot taken to be possibly actual, that it cannot be believed to be actual by someone who has not already made the leap of faith making him a skeptic. The skeptic as I have imagined him produces a skeptical scenario in which you would not know what you take yourself to know and thereby challenges your knowledge claims. Simply to show that skepticism is an uncomfortable or perhaps even impossible theoretical option does not meet this challenge, for the doubts induced by the skeptical argument may keep nagging you. An adequate answer to skepticism therefore has to be specific, explaining in detail why the skeptical scenario is not a doxastic alternative *for us*.

The local anti-symmetry of the doxastic alternativeness relation, to which belief in either (5.1), (5.6) or (5.7) commits me to, underlies the pragmatic indefensibility of these claims. Suppose, again, that *a*, in *w*, believes what *a* would express by uttering "*p*, but I do not believe that *p*." If his beliefs are consistent, then there is a doxastic alternative *v* which *a* believes to be a way the actual world might be where both conjuncts of this sentence are true. In this doxastic alternative, *a* would have other beliefs than he actually has or else would have inconsistent beliefs. Does this allow him to rule out *v* as a way the actual world might be? Well, it depends. Under a externalist reading of "the actual world", he may very well imagine it to be different with respect to the beliefs he has. Under an

---

40. Suppose $w \models p \wedge \textbf{Bel} \neg \textbf{Dis} \neg p$. Suppose there is a v such that $w \mathscr{DA} v$ and $v \mathscr{DA} w$. Then $v \models \neg \textbf{Dis} \neg p$, that is $w \not\models p$ which lands us in a contradiction.

internalist reading, however, this is not possible. For if it were actual, he would not take it to be actual, for he would, presupposing consistency, have different beliefs about it than he actually has. In the case of (5.7), we may strengthen this point, for we do not have to rely on the dubious assumption of consistency, which begs the question the skeptic wants us to consider. (5.7), then, is indefensible because belief in it can only be true if we only take worlds to be doxastic alternatives which we recognise to be worlds which might not be actual.

It is this commitment to a locally anti-symmetric alternativeness relation that justifies our reluctance to believe, even if it is for the sake of the skeptic's argument, that we do not disbelieve that his counterfactual skeptical scenario is actual. What the skeptic asks us to do, in effect, is to suspend our confidence in our knowledge claims, i.e. to accept the following as true:

(5.8) I know that *p*, but I believe (for the sake of the argument) that I do not disbelieve that I do not know that *p*.

If the second conjunct is true and I have the belief that I am lacking the confidence I actually have in my knowledge claims, then the first conjunct is true only in worlds which are not doxastic alternatives to any of my doxastic alternatives. Thus (5.8) asks us to epistemically place ourselves in worlds from which there is no road back to the world where we take ourselves to be. By the first conjunct, the suspension of our normal confidence in our knowledge claims is only problematic for someone who actually endorses them. This is why (5.8) is a problem for the non-skeptic, but not for someone who is a skeptic already, i.e. who has taken the leap of faith required for this conversion. .

If skepticism requires a leap of faith we have good reasons not to take, we treat the skeptic fair in disbelieving him without proving him wrong.

5.1.5 Meeting the skeptical challenge

One particularly popular way of facing the question whether we know something is by asking how we know that we are not dreaming now. If we do not know that we are not dreaming now, how then can we be sure that we not only *seem* to know something? Here is Descartes asking himself this question:

> "Mais, en y pensant soigneusement, je me ressouviens d'avoir esté souvent trompé, lors que je dormais, par de semblables illusions. Et m'arrestant sur cette pensée, je voy si manifestement qu'il n'y a point d'indices concluans, ny de marques, assez certaines par où j'en puisse distinguer nettement la veille d'avec le sommeil, que j'en suis tout etonné; & mon estonnement est tel, qu'il est presque capable de me persuader que je dors." (AT IX/I 15)

It is important that Descartes does not have to *suppose* that he is dreaming in order to challenge the knowledge claims based on basic perceptual beliefs like the one that he is sitting by the fire, etc. It is enough if he becomes awake to the possibility that he *might* be dreaming while keeping his beliefs. If he recognizes this to be possible, he has to treat a world where he only seems to know what he takes himself to know as a doxastic possibility. This undermines his knowledge claims by forcing him to acknowledge that they might be false. So he does not know them to be true and the skeptic has won.

Recently, Descartes' dream argument has been generalized in interesting ways. I suppose the story is familiar by now: there is this uncomfortable place out there in the realm of possibilities where you are a brain in a vat, floating in a nutrient fluid keeping you alive, with a clever superscientist stimulating your brain to have exactly the same as (or, at least, introspectively indistinguishable from) your present experiences. By hypothesis, everything you experience in that situation is exactly like what you are experiencing now, with the sole difference that there you do not have any clue what the external world is like. So how can you be sure that the matrix does not already have you now?

In fact, you cannot. What I want to show is something much more modest: *If* you do not already believe that you might be a brain in a vat, it is impossible to convince you that this scenario really is a possibility you might be in. Suppose you are dealing with a skeptic who tries to convince you that you might be a brain in a vat. I am able to vaccinate you against becoming a skeptic iff I manage to convince you of the truth of what you would express by

the following:

(5.9)
"If I were a brain in a vat, I wouldn't take the actual world for what I actually take it to be, namely possibly actual."

(5.9) would mean that the skeptical scenario exhibits the structure of (5.7), committing us to anti-symmetric doxastic alternatives: it cannot be believed to be true in a world one takes oneself to believe to be actual. It seems that we cannot be brains in a vat unless we disbelieve that we are not. If this were the case, then the skeptic could not hope to bring us to suspend our disbelief: if he were to succeed in this, it would follow from our failure to disbelieve that we are not brains in a vat that we actually are not. So he would not have convinced us. To dispell this doubt, the skeptic has to convince us that we might be both brains in a vat and not disbelieve that we are not. He has to make his scenario possibly actual for us. Given that it exhibits the structure of (5.7), this cannot be done.[41] My argument for (5.9) rests on a crucial property of both the dream- and the brain-in-a-vat-scenario: If you were dreaming now, you would not believe you were dreaming. If you were a brain in a vat, you would not believe you were a brain in a vat. For suppose otherwise. Then you would not have any reason to worry about these scenarios being actual. For if you would think you are a brain in a vat provided that you were one, then there would be no way of convincing you that you might be a brain in a vat, given that you do not believe it already. For if you do not believe it already, you thereby differ from a brain in a vat and so you can, following the argument just sketched, be sure that you are not in such a deplorable situation. What I claim here, in effect, is that someone trying to believe the skeptic finds himself committed to believe something having the form of Moore's paradox (5.1), i.e. "I am a brain in a vat but I do not believe I am a brain in a vat." A skeptic, however will not be in this state, for he already believes that he might be a brain in a vat and therefore does not differ from a brain in a vat which believes that it is a brain in a vat.

We can go further, however, and show that belief in the *possibility* even of the skeptical scenario requires something of the form of (5.7), namely "I am not a brain in a vat but I do not disbelieve that I am a brain in a vat". This would mean that the doxastic alternative in which I am a brain in a vat does not have the actual world, where I am not a brain in a vat, as a doxastic alternative. Considering it possible commits us to an anti-symmetric alternativeness relation: we loose the way back to the actual world.

The argument for (5.9) now is the following: Suppose you are a brain in a vat. Then you do not believe that you are a brain in a vat. So there is a doxastic alternative where you fail to be one. This doxastic alternative, however, is not the one you suppose to be in. So it is one you would not – under this supposition – take to be actual. The doxastic alternative you would in this case take to be actual is one where you *are* a brain in a vat, so you would, if you are a brain in a vat, disbelieve that you are not a brain in a vat. Supposing otherwise would mean to describe the scenario as one in which you fail to be a brain in a vat in all doxastic alternatives (to that scenario), i.e. to suppose that the (under the supposition) actual world where you are a brain in a vat would not be a doxastic alternative. But you take, even under the supposition that you are a brain in a vat, the actual world to be actual. So if you were a brain in a vat, you would not take the actual world to be among your doxastic alternatives, i.e. you would not take it to be possibly actual.

The skeptic who wants to counter this argument has to argue that even if we were brains in a vat, we would not disbelieve that we are not. That is, he has to argue for (5.7) which, as we saw, we cannot become convinced of.

The crucial thing about (5.9) is that it only gives us a reason to disbelieve the skeptic if we take "actual world" to refer to the world which *we*, non-skeptics, take to be actual. In this case, however, it gives us what it takes to inoculate us against the skeptical disease. Here then is another version of the argument:

(1) Suppose that there is a world *w* where I am a brain in a vat.
(2) I therefore do not, in *w*, believe that I am a brain in a vat.

---

41. To see this, note that (5.9) follows from $p \rightarrow D\neg p$ (where p is "I am a brain in the vat"): if the brain-in-a-vat world were actual, I would, disbelieving that I am not a brain in a vat, reject the real actual world even as a doxastic alternative. This means that convincing us of the possible actuality of the brain in a vat scenario would require to dispel that impression, by bringing us to believe that $p \wedge \neg D\neg p$, i.e. to make us believe (5.7).

(3) So I have, in *w*, a doxastic alternative which does not make it true that I am a brain in a vat.

(4) Given that, in *w*, I would be supposing that I am a brain in a vat in *w*, this world cannot be *w*.

(5) If I would not disbelieve in *w* that I am a brain in a vat, no world where I am a brain in the vat would be among my doxastic alternatives.

(6) In particular, *w* itself would not be among my doxastic alternatives: *w* would be a non-reflexive world.

(7) But I cannot take a non-reflexive world to be possibly actual.

(8) So I disbelieve, in *w*, that I am not a brain in a vat.

(9) So (discharging the supposition): Supposing that there is a world *w* where I am not a brain in a vat, I have to ascribe to myself, in *w*, disbelief that I am not a brain in a vat.

(10) So I would not, in such a world, believe that I am not a brain in a vat.

(11) But I believe that I am not a brain in a vat.

(12) So I am not a brain in a vat.

This argument shows that you cannot suppose that you are a brain in a vat without having to concede that, actually, you are not a brain in a vat. You cannot consider the brain in a vat scenario to be possibly actual. The crucial premiss (11) makes it impossible to use the argument outlined above to convince the skeptic. For not believing that ¬*p* is, in general, compatible with there being a doxastic alternative where ¬*p*, as long as there is another doxastic alternative where *p*. For the skeptic, the existence of a doxastic alternative where he is a brain in a vat is guaranteed. For us, however, it is precisely this existence that is at stake. Given that we already believe that we are not brains in a vat, no such doxastic alternative exists. So the skeptic's convinces only those who are believers already. It requires a leap of faith.

5.2 Epistemic normativity

5.2.1 Epistemic oughts are really mights

5.2.2 Epistemic supererogation

5.2.3 Apriority as an evaluative notion

5.2.4 Against the contingent a priori

5.2.5 Epistemic subrogatives

5.3 An Epistemology of Attitudes

5.3.1 Affective knowledge of values

Cognitivists accept

(5.10)

    x regrets that *p* → (*x* correctly regrets that *p* iff it is regrettable that *p*)

"Friends of correctness conditions" (Rabinowicz & Rønnow-Rasmussen, Skorupski, Scanlon, Mulligan) subscribe to

(5.11) If *x* correctly regrets that *p*, then (*x* correctly regrets that *p* because it is regrettable that *p*)

while buck-passers or neo-sentimentalists combine (1) with

(5.12)

    If it is regrettable that *p*, then (it is regrettable that *p* because (*x* regrets that *p* → *x* correctly regrets that *p*))

The problem of buck-passing theories of value is that it gets the explanatory direction wrong: the Evil Demon who inflicts pain unless I prefer the saucer of mud makes my preferring it valuable, but does not make eating the saucer of mud valuable.

They also cannot explain why affective knowledge of value can motivate:

> "Suppose with the neo-sentimentalist that being valuable is understood in terms of appropriate emotions or good, undefeated reasons to feel emotions. What, then, would knowledge of the value of an object amount to? […] presumably a neo-sentimentalist must hold that we sometimes have knowledge of the appropriateness of an emotion. And in the most basic cases this could only be knowledge that an emotion is appropriate, that there are undefeated reasons to fteel an emotion. But if we have knowledge of values it is extremely implausible to think that such knowledge consists only of knowledge that, a knowledge by description which has no anchorage in any knowledge by acquaintance. On one common and plausible view, knowledge that p cannot motivate; even axiological knowledge, knowledge that it is valuable that p, is an intellectual state and, like all such states, cannot motivate. Feeling (dis)values, however, is no intellectual state and can motivate." (Mulligan 2007: 225)

On Mulligan's alternative account, emotions are reactions to feelings of value:

> "Suppose that Maria is walking down the street and observes a scene in which bread is being distributed unequally to equally needy children. She is struck, as we say, by the injustice of the situation. She has felt the injustice of the situation[.] Perhaps she reacts with indignation. Perhaps she is suffering from indignation fatigue and feels no emotion whatsoever. " (Mulligan 2007: 223)

Because emotions are reactions to feelings of values, they not only have motivational force but may also be justified by perception. My fear of the dog is justified not just by its dangerousness (which is circular if its dangerousness consists in fear being appropriate), but also by my sensation of its dangerousness, which is not a representation of danger, but rather a perception of the dog as dangerous.

Problem 3: What is seen in the perception of the dog as dangerous? **Adverbialist solution**: we see the picture as beautiful, the dog as dangerous, the act as shameful.

Against prescriptivism: the normativity of mental acts is not a matter of them being obligatory, but exclusively of them being permissible - as the coherence, and apparent blamelessness, of both the sceptic and the stoic position show (mental states, like perception, are different; they are obligatory under certain circumstances):

the sceptic acknowledges things as true, but does not react with belief – stays with appearances;
the stoic acknowledges things as offensive/dangerous etc. but does not react with anger/fear etc – stays with feelings of values (cannot avoid being struck), but does not respond with an emotion that would be appropriate

Distinction between having beliefs and emotions AT ALL – there may very well be conditional obligations even for beliefs and emotions (being afraid of, being relieved)

My interpretation of the cases: they still get the form (ie they perceive), but they don't react

5.3.2 Resisting counterevaluatives

ABSTRACT In my talk, I discuss the peculiar behaviour of counterevaluatives, sentences that claim evaluative matters to be different from how they actually are. Counterevaluatives have recently received a lot of attention in discussions of the so-called "puzzle of imaginative resistance" (Gendler 2000). Distinguishing between quasi- and make-believe mental states, I trace the difficulties we experience in imagining counterevaluatives to the way in which counterevaluatives involve quasi-emotions and quasi-feelings. I'll argue that both the fearfulness of the dangerous and our incapability to imagine it being otherwise are consequences of the way emotive states in general behave under off-line simulation: it is because the relation between quasi-fear and real fear is different from that between supposition (make-believe assertion) and assertion that we cannot imagine not being afraid of the fearful.

The constitutive relation between emotive quasi-states and the formal objects of their real counterparts should also shed some light on normative necessity. Emotions are said to involve formal objects, in the way fear involves the fearful (Kenny 2003: 134). It is with respect to their formal objects that an emotion's appropriateness may be assessed (cf. Goldie 2000: 34) . While they clearly play an explanatory rôle both in our every-day practice with and our theoretising about emotions, the ontology, metaphysics and epistemology of emotions' formal objects remain largely unexplored. In my talk, I would like to make some progress on this score, focussing on their modal behaviour: in what sense can it be said that the fearful could have failed to make fear appropriate? in what sense can we imagine not fearing the fearful? Taking a lead from Tamar Gendler's "puzzle of imaginative resistance" (Gendler 2000), I'll argue that both the fearfulness of the fearful and our incapability to imagine it being otherwise are consequences of the way emotive states in general behave under off-line simulation: it is because the relation between make-belief fear and real fear is different from that between supposition and assertion that we cannot imagine not being afraid of the fearful. This does not, however, mean that it is necessary that we fear the fearful.

The puzzle, a range of solutions, and the connection with emotions

David Hume writes in *On the Standards of Taste*:

> "Where *speculative* errors may be found in the polite writings of any age or country, they detract but little from the value of those compositions. There needs to be but a certain turn of thought or imagination to make us enter into all the opinions which then prevailed and relish the sentiments or conclusions derived from them. But a very violent effort is requisite to change our judgment of manners, and excite sentiments of approbation or blame, love or hatred, different from those to which the mind from long custom has been familiarized […] I cannot, nor is it proper that I should, enter into such [vicious] sentiments."

Though the exact interpretation of this passage is controversial, Hume distinguishes between the reactions we have towards assertions that contradict our descriptive and evaluative beliefs respectively. At least part of the puzzle this raises can also be illustrated by the following:

> Try to imagine the following fictional mini-story to be true:
> "In killing her baby, Giselda did the right thing; after all, it was a girl.â[42]

When faced with such an imperative, we face what we may call "imaginative resistance". If this is accepted as a fact, then the following puzzle arises:

> "…the puzzle of explaining our comparative difficulty in imagining fictional worlds that we take to be morally deviant." (Gendler 2000: 56)

For the sake of exposition let us indulge in the fiction that we imagine worlds, possible or impossible, and let us call the world we are to imagine the Giselda-world. In the following, I am going to assume that we have indeed comparative difficulty in imagining morally deviant fictional worlds and will discuss several explanations philosophers have given of this fact, before offering my own.

Let us first note, however, how vague our starting point really is. Gendler frames it in terms of an asymmetry in make-belief:

> "I cannot bring myself to believe that murder is right – but I cannot bring myself to believe that the earth is flat either. When it comes to make-belief, however, we seem more inclined to find ourselves stumped in the one case than in the other." (Gendler 2000: 58)

In a footnote, she then notes:

---

42. (Cf. Walton 1994: 37). I follow Gendler (2000: 56) in taking this to be a – somewhat contrived – example of an alleged fictional truth evincing imaginative resistance, ignoring its possible literary contexts.

"Walton points out (personal correspondence) that my use of 'make-believing' seems ambiguous between two readings. If I make-believe that *p*, I may be: (a) accepting that *p* as been successfully made fictional (that is, accepting that the author has succeeded in presenting a story in the context of which a certain proposition is true) or (b) pretending that *p* (that is, entertaining or attending to or considering the content of *p*, in the distinctive way required by imagination)." (Gendler 2000: 58, fn. 6)

Weatherson (2004) has called (a) the *alethic* and (b) the *imaginative* puzzle of imaginative resistance. In the following, I will concentrate on (b), though I take our answer to the (b)-problem to have implications on (a) as well – at least if works of fiction are invitations to imagine (Walton 1990). I do not claim that we never succeed in accepting morally deviant propositions to be fictionally true – context and rhetoric may help us to do so. I think, however, that there is a comparative – not necessarily an absolute – difficulty in imagining morally deviant worlds and that this explains why context and rhetoric are required for this.

A range of solutions

In recent philosophical discussions, imaginative resistance has been explained by:

  (i)   our refusal to imagine (Gendler);
  (ii) our difficulty of imagining (Hume, Moran, Currie); or
  (iii) our inability to imagine certain propositions (Walton, Weatherson, Yablo).

The argument of Gendler (2000) consisted of two parts: she first argued that our difficulty in imagining the Giselda-world is not due to its being logically or conceptually impossible, because we succeed in imagining (at least some) logically or conceptually impossible worlds and because we face imaginative resistance also with respects to worlds that are clearly logically and conceptually possible. Though I am not very much convinced of her first reason, the second – which I take to be sound – suffices to discard conceptual impossibility as the only source of imaginative resistance. In a second part, Gendler sketched a positive account, turning on the alleged fact that the reader of fictions "feels being asked to export a way of looking at the actual world which she does not wish to add to her conceptual repertoire" (Gendler 2000: 77).

Unfortunately, Gendler's positive proposal is rather vague – unwillingness to add a certain perspective to one's 'conceptual repertoire' can explain unwillingness to imagine the Giselda-world only if it is properly distinguished both from *believing* that murder is right and from *supposing* that it is. The assumption that the author of the Giselda-story presumably wanted us to convince that murder is right is gratuituous – we are not required to make it in order to experience imaginative resistance. The order of explanation is rather in reverse: it is because we experience imaginative resistance that we are unwilling to accept the author's invitation to change our moral beliefs if really such an invitation is made. But in the same way, we are unwilling to change our non-moral belief if some author would want to convince us that really, pigs can fly.

The impossibility hypothesis, as Gendler (2000: 64) calls it – i.e. the view that we cannot imagine the Giselda-world because it is conceptually impossible – is just one way of defending an inability-diagnosis of imaginative resistance. A more general diagnosis is that imaginative the Giselda-worlds goes against our belief in certain dependence ties:

"Our reluctance to allow moral principles we disagree with to be fictional is just an instance of a more general point concerning dependence relations of a certain kind." (Walton 1994: 46)

The dependence relations here are those holding between the subvening base of non-moral properties and the moral properties supervening on them in the sense that no two things could differ in moral properties without also differing in non-moral ones. Weatherson (2004: ??) has argued that we experience imaginative resistance in all and only the cases where the following supervenience principle is violated (he calls it "virtue"):

"If *p* is the kind of claim that, if true, must be true in virtue of lower-level facts, and if the story is about those lower-level facts, then it must be true in the story that there is some true proposition *r* which is about those lower-level facts such that *p* is true in virtue of *r*." (Weatherson 2004: 18)

Weatherson (2004) claims that radically divergent epistemic evaluations and attributions of mental states and content also violate this principle and are therefore difficult to imagine.

The problem, however, is that our belief in supervenience does not account for our "inability to understand fully what it would be like for [the moral facts] to be different" (Walton 1994: 46), or, as Gendler (2000: 66) phrases it, of why "we cannot make sense of what it would be for something to be both an instance of murder and an instance of something that is morally right, and for it to be morally right *because* it is an instance of murder". For our belief that some supervenience relation holds is itself descriptive and therefore readily available for suspension in imagination. But it is not easier to imagine the truth of

> "In a world where the total distribution of physical qualities over space-time is importantly different from how it is in the actual world, Giselda did the right thing in killing her baby; after all, it was a girl."

But supervenience-violation is neither necessary nor sufficient for imaginative resistance.[43] It is not necessary for we experience imaginative resistance also in cases where no supervenience relations are violated, e.g. where the lower-level facts are imagined to be radically different from what they are.

It is not sufficient because many violations of supervenience-relations are readily imagined. We imagine, for example, that one morning, Gregor Samsa, waking up from anxious dreams, discovered that in bed he had been changed into a monstrous verminous bug – *even if* we think that Gregor Samsa is essentially a person. This is even clearer in the case of entities of other categories: their spatio-temporal location is plausibly essential to events, but we readily imagine my talk having taken place 40 minutes earlier than in fact it actually does.

The connection with emotions

Stephen (Yablo 2002: 485) argues that we have difficulty in accepting a story where maple-leafs are oval. He generalises this to grokking or response-enabled concepts. A response-enabled concept like "oval" is picked out, but not analysable in terms of our perceptual responses: while we pick out oval things as those that look roughly egg-shaped, it is not necessary that oval things look egg-shaped – even in a world where perceivers have radically different perceptual mechanism than we do, there still would be oval things. The question whether some things in a possible world are oval is not to be decided by how they look to their world-mates, but by how they look to us.

Currie (2002: 205) draws an interesting contrast between belief-like and desire-like imagining. Belief-like imagining is supposing – what we do if we evaluate indicative conditionals by the Ramsey test. In belief-like imagining, the inferential patterns of belief contents are preserved, even though imagined propositions are not required to be consistent with each other and with our beliefs. Desire-like imaginings, according to Currie (2002: 210–211) occur when "we imagine ourselves in [a] situation and then, in imagination, we decide to do something". The desire of the theatre-goer that Othello not kill Desdemona is of this kind – and it is not the desire that, in this fiction, Desdemona not be murdered. The latter is a real (but odd) desire, the first a desire-in-imagination (Currie 2002: 212). It is in conjunction with desire-like imaginings that belief-like imaginings have emotional consequences. Imaginative resistance, according to Currie, arises when we choose to preserve the normal connection between belief and desire (and between belief-like and desire-like imagining) at the cost of desire-like imagining:

> "…if it is difficult [for the reader] to have the desire-like imagining that female infants be killed, she can have the belief-like imagining that female infanticide is right only at the expense of the harmony between belief-like and desire-like imagining which is the natural stance of the intelligent and sensitive reader." (Currie 2002: 217)

As Currie (2002: 218–219) acknowledges, this diagnosis does not cover all cases: it does not explain our difficulty in imagining the lame joke to be funny or imagine some pain we would experience if tortured by gentle contact with a fluffy chair.

"it is harder, much harder, to get people to desire in imagination against the trend of their own real desires than it is to get people to believe in imagination against the trend of what they really believe" (Currie 2002: 214)

---

43. I am indebted here and in the following to Tyler Dogget's talk on imaginative resistance at MIT in spring 2005.

Gendler characterises the cases of imaginative resistance Hume had in mind as "cases involving valenced normative evaluations: we are asked to assess something as mannerly or unmannerly, praiseworthy or blameworthy, loveable or hateable, where each of the pairs identifies two points along a normative spectrum where one end is desirable and the other is not".

Gendler

Criticisms of Gendler
****

Currie, Gregory (2002). âDesire in Imaginationâ, in Gendler and Hawthorne (2002), pp. 201-221. Doggett, Tyler (2004). âMoral Properties and Moral Imagination.â Unpublished PhD Dissertation, MIT. Holbo, John (2003a). âImaginative Resistance.â At http://homepage.mac.com/jholbo/homepage/pages/blog/giantHolbo, John (2003b). âThe Varieties of Imaginatively Resistant Experience.â At http://homepage.mac.com/jholbo/homepage/pages/blog/ Derek (2003). âFictional Assent and the (So-Called) âPuzzle of Imaginative Resistance.ââ In Kieran and Lopes (2003). Kieran, Matthew and Dominic M. Lopes, eds. (2003) Imagination, Philosophy and the Arts (London: Routledge, 2003), Mothersill, Mary (2002). âMake-Believe Morality and Fictional Worldsâ in Bermudez and Gardiner (2002). Bermudez, JosÃ© Luis and Sebastian Gardiner, eds. (2002). Art and Morality. London: Routledge. Stokes, Dustin (forthcoming). [Title] Stock, Kathleen (2003). âThe Tower of Goldbach and Other Impossible Tales.â In Kieran and Lopes (2003), 107-124. Kieran, Matthew and Dominic M. Lopes, eds. (2003) Imagination, Philosophy and the Arts (London: Routledge, 2003), Stock, Kathleen (2005). ââResisting Imaginative Resistanceâ Philosophical Quarterly. Weatherson, Brian (2004). âMorality, Fiction and Possibility.â Philosophersâ Imprint. Volume 4: Number 3. <www.philosophersimprint.org/004003/> Yablo, Stephen (2002). âCoulda, Woulda, Shouldaâ, in Gendler and Hawthorne (2002), pp. 441-492.

Gendler, the puzzle of imaginative resistance Kevin had a plausible and minimalist explanation of imaginative resistance (he denies the phenomenon, but this is another thing): evaluative judgments are tied up with feelings; feelings (i) cannot be simulated or (ii) can only be simulated in rather special ways (as some kind of make-believe feelings). (ii) would explain the doubling of the narrator. Compare this with the imaginative resistance of a vegetarian to "if you haven't had a good steak for days, you just feel terribly hungry".

This is related to the – interesting, and difficult – question whether we experience real fear say in movie theatres. If fear is intrinsically unpleasant, I would rather say no - it's a kind of make-believe fear, not only fear based on false presumptions (does not have to be), but not real fear. If fear may not be unpleasant (eg for masochists), then what kind of fear is it? (masochism alone cannot explain popularity of horror films)

Make-believe in general

We need, in other words, a logic of "make-believedly" that applies to emotive and not just conative actions.[44]

An alternative to the voluntarist explanation:

1. Evaluative judgments are tied up with feelings.
2. Feelings either cannot be simulated or they can be simulated only in rather special ways (as some kind of make-believe feelings, as perhaps they are when we go to the movies).

Compare this with the imaginative resistance of a vegetarian to "if you haven't had a good steak for days, you just feel terribly hungry".

(1) Suppose that murder is right.
(2) Imagine that murder is right.
(3) Accept as fictionally true that murder is right.
(4) Pretend that murder is right.

---

44. It should apply to them 'directly', not just in a round-about way via a detour through fictional truths that include "such propositions as that one is now experiencing fear and pity for the tragic hero, or that one feels the satisfactions of vengeance when rough justice is meted out to the deserving" (Moran 1994: 77). Imagining that the proposition "I am feeling fear of the lion" is true (or imagining it as true), however, is clearly different from make-believedly fear the lion: feeling fear-at-the-movies is not an activity involving propositions.

(5) Make-believedly approve of murder.

(1), even by Gendler's lights, is easy. (2) is what she takes to be difficult, understanding it as (at least entailing) (3). She takes the difficulty of (4) to explain the difficulty of (3). I think the difficulty (if any) of (2) is explained by the difficulty of (5).

How wide is the phenomenon of imaginative resistance? A lot of cases have been mentioned in the literature:

  (i) We have difficulty in imagining, of a really lame joke, that it is funny (Walton 1994: 43–44).
 (ii) We have difficulty in imagining, of sour milk, that it smells good (Gendler 2000: 78 attributes it to Carl Ginet).
(iii) We have difficulty in imagining, of a monster truck rally, that is is sublimely beautiful (Yablo 2002: 485).

The following are easy:

(E1) Suppose that this tomato looks green.
(E2) Suppose that this joke is funny.
(E3) Suppose that sour milk smells good.
(E4) Suppose that you are a zombie.
(E5) Suppose that murder is right.

But the following are difficult:

(D1) Make-believedly see this tomato as green.
(D2) Make-believedly laugh about this joke.
(D3) Make-believedly like this smell.
(D4) Make-believedly be all dark inside.
(D5) Make-believedly approve of this murder.

Distinguish

  1. fear, understanding, assertion, having as formal objects the dangerous / meaning (?) / truth.
  2. quasi-fear (in movie theaters), quasi-understanding (e.g. empathy), quasi-assertion (on stage); same psychological feel, but different connection with action and reasoning.
  3. make-believe / as-if fear, understanding, assertion (supposition): no or different psychological feel

You can make-believedly understand "kdflkdslf" if reading a story in which you do, but you do not quasi-understand it. You can make-believe fear and suppose fearing it, but you cannot quasi-fear something you do not at the same time take to be dangerous.

Compare

  1. He said that 'philtosophers' stink.
  2. He said that 'niggers' stink.

Are scare quotes monsters?

what explains this comparative difficulty?

essential properties of the objects of emotions

restricted counterpart relations – de re

Teroni - 2 notions of 'formal object':

  1. what gives the relevant act its conditions of satisfaction
  2. what makes the specific act the act it is

cf Frege on the grasping of truth-values – judging is marking out a part within the True – this is ridiculous, but the corresponding claim for practical attitudes seems plausible: to fear a spider, is indeed to mark out part of the dangerous as an object of an attitude

But as a de re claim (accommodating KM's point) Even if the fear of a spider is not the fear of the dangerous, the fear of a spider is a fear of the spider as dangerous

– but of things as so-and-so –> FORMAL OBJECTS ALSO FOR FEELINGS, CONTRA MULLIGAN

This is a de re claim: not only is fear fear of something perceived as dangerous but this state of fear is essentially of something perceived as fearful

including Feeling because I include finding funny – I suggest that she may get the point of the joke, feel its funniness and yet not respond emotionally in any way at all.

upshot for normative necessity: pleasure is good, pain is bad, the beautiful is admirable or knowledge is better than error etc - it is not always de dicto - it may be metaphysical necessity after all, holding in virtue of the essences of the objects of emotions

It also accounts for the fact that some readers, e.g. Gendler (2000: 74), react to the Giselda-story by "doubling of the narrator", i.e. concluding not that in the story, murder is right, but that *the narrator thinks* that, in the story, murder is right, and explains why we often consider the morally deviant authorial comments as redundant.

normative necessity

Evidence for the claim that the analogy is normally made

Analogy thesis committed to the claim that the difference between watching something which is believed to be a documentary is only gradually different from watching the same thing and believing it to be fictional.

"as if" can mean "others, not me" or "false, not true"

is "off-line" rather a mode or an operator?

EMPATHY terminology: one species of empathy is mimicry, which sub-divides into imitation and feedback

defined by Hoffman (2000) as "vicarious affective response to another person" MUCH TOO BROAD it is automatic and probably even hard-wired -> Hoffman (2000: 44), Italian neurologist met in Bologna "mirror-neurons" it is noninferential: Wittg Zettel §537 is it direct: Eros: no: "I see Mary's pain" is ungrammatical; Mulligan: yes, seeing sadness

what it is not - supposition - imagining something to be fictionally true - taking something to be true

authorial authority: there are some limits to what authors can make fictionally true, eg. no-one can make the sentence "This sentence is not fictionally true" fictionally true.

wie verwendest Du "mbly" in "x make-believedly sees"? Walton verwendet doch "make-believe" ziemlich wÃ¶rtlich, dh als "make s.o. believe", dh jemanden was glauben machen. Du scheinst es eher wie "off-line" zu verwenden. Aber was ist off-line sehen, off-line admire usw.? Einige MÃ¶glichkeiten: x make-believedly sees y = - x tut so, als sÃ¤he er y - x macht (z) glauben, dass er y sieht - x stellt sich vor, dass er y sieht - x sieht* y, wobei Sehen* so zu Sehen steht wie Kinoangst zu Angst - x "indulges in the fiction that" er y sieht Mir scheinen alle 5 verschieden.

Imaginative resistance – it's nature and scope

Gendler (2000: 77) a tenté d'expliquer cette résistance par le fait que l'on demande alors au sujet de modifier son répertoire conceptuel de façon à y incorporer une manière d'appréhender le monde actuel qui lui répugne.

Indépendamment de la valeur de cette explication,[45]

Gendler a certainement raison de souligner l'impact global de ce genre de suppositions : une supposition allant à l'encontre de nos convictions morales, ou plus généralement de nos croyances évaluatives, ne se cantonne effectivement pas à l'intérieur du scénario imaginé, mais empiète sur ce que nous croyons réellement.

Le phénomène ne se limite pas à la sphère morale, mais concerne toutes les évaluations: nous éprouvons de la difficulté à imaginer qu'une blague plate soit hilarante (Walton 1994: 43–44), que du lait périmé soit délicieux

---

45. Je ne suis pas convaincu par ce diagnostique, qui me semble commettre l'erreur d'attribuer tout effet possible d'une fiction à son auteur. L'ordre de l'explication est plutôt inverse: s'il y a une telle intention de l'auteur, c'est parce que nous ressentons une résistance imaginative que nous refusons l'invitation d'entrer en jeu.

(Gendler 2000: 78 attribue cet exemple à Carl Ginet), ou encore qu'une rallye de 30 tonnes soit sublime (Yablo 2002: 285).

Dans tous ces cas, cependant, deux autres activités restent possibles. Nous pouvons nous imaginer qu'une blague plate soit *considérée* comme hilarante, que le lait périmé soit *apprécié* ou que la rallye soit *jugée* sublime. De même, nous pouvons nous imaginer qu'une autre blague que celle que nous avons entendue soit hilarante, que le lait seulement légèrement périmé soit délicieux et qu'une rallye en forme de ballet soit sublime. Dans ces exercices d'imagination, nous changeons soit l'attitude, soit l'objet. Ce que nous ne pouvons pas faire, cependant (ou, au moins, avons de la difficulté à faire) est de rompre, même par l'imagination, le lien étroit qui est entre l'objet et notre réaction évaluative.

5.3.3 Intentional objects

Laurent Cesalli: INTENTIONAL OBJETS IN MARTY (AND BRENTANO) intentional inexistence is existence-in, not non-existence the intentional relation is between the mental act and the immanent (represented) object and also between something that is real and something that is not (ie the immanent object is NOT real) two ways of making the real/immanent object distinction: (i) one thing existing in two ways; (ii) two things existing simpliciter. In 'Raum und Zeit' (1916, posthumously published, he scraps the immanent object and seems to go disjunctivist: instead of similarity between the real and the immanent object, we have similarity between the mental act and its object – IN CASES WHERE THERE IS SUCH AN OBJECT. A possibilium is something that exists without being actual (real?) INTENTIONAL OBJECTS IN OCKHAM: the mind creates a fiction, a fictum that has only objective reality; he restricts universals to such immanent objects. PB DOES OCKHAM ALSO THINK THAT TO EXIST IS *TO BE* SINGULAR? THAT'S PLAUSIBLE. Later, he scrapped the ficta, and identified the concepts with the intellectual act itself!! (like Marty): the mental act can be similar to the object, can signify and replace the external object

Oskar Kraus has a biography on Anton Marty Preissschrift on Su. Th. I q85, art 1-3 Würzburg 1868 - Brentano (who had 30 years, while Marty had 21) 1875 Über den Ursprung der Sprache - Habilitation in Göttingen teaches in Prag Max Brod and Franz Kafka

INTENTIONAL OBJECTS One of the thing Crane thinks intentional objects are important for is that it allows us to rebut propositionalism. But if all intentional states have truth/satisfaction conditions, and granting the point that these conditions will be propositional, then all intentional states are propositional. Moreover, Crane does think that necessarily, content = propositional content. If I fear the dog down the road, then, presumably, the dog is presented to me in a certain way, say as fearful. But that does not mean that, thereby, my intentional act of fearing the dog has conditions of satisfaction. So, aspectual shape and content cannot be reduced to conditions of satisfaction. These are two different ideas. The debate between Gorman and Crane therefore does not boil down to what the other thinks it does. GORMAN SAYS: Gorman?s worry about Crane amounts to this: Crane explains ?being an intentional object? as if it were a role played by substantial objects?a boat, for example, is a substantial object, and when it is thought about, it (the boat) plays the role of an intentional object, a role that doesn?t depend on its (the boat?s) having any particular substantial nature. But Crane does not explain how this role?being an intentional object?can be played when there is no substantial object at all. In other words, Crane does not explain how the schematic notion of an object can be applied to the precise case at issue, viz., when there is no substantial object.

INTENTIONALITY

> Any philosophy of intentionality, of the property peculiar to mental acts, states and activities of being "directed" towards or about something, should contain many chapters. It should provide an account of the different mental acts, states and activities. It should provide an analysis of the relations and other ties hiding behind the metaphor of directedness. And it should provide an account of the sorts of things mental acts, states and activities are directed towards. A philosophy of intentionality should, further, tell us about the intentionality of all the main types of mental states, acts and activities. It should tell us, at the very least, about the intentionality of

> acquaintance, admiration, attention, belief that, belief in, certainty, choice, deliberation, desire, doubt, expectation, hate, hope, imagination, judgement, knowledge, love, meaning that p, memory, perception, preference, regret, shame, sympathy, striving, supposition, time- consciousness, trust, uncertainty, understanding, vision, willing and wishes

and not limit itself to, say, the intentionality of belief and desire. (Mulligan 2007: 205)

DISTINGUISH perceptual objects FROM objects of perception

"CONSTRUCTION" OF THE (MULTISENSORY) PERCEPTUAL OBJECT – this solves Molyneux per fiat, cf "I have argued elsewhere that some multisensory effects, revealed by cross-modal illusions, resolve conflicts across modalities. Conflict requires a common subject matter, so performing conflict resolution demonstrates a shared perceptual concern for the common sources of stimulation to multiple senses (O'Callaghan 2012)." (O'Callaghan 2016: 1271) – Several philosophers have posited sensory individuals as perceptible feature bearers (e.g., Strawson 1959; Clark 2000; Cohen 2004; Nanay 2013, §3.3). OBJECT OF PERCEPTION: O'Callaghan (2016: 1274) even understanding "bodies" so as to include rainbows, holes, flashes of light is still too restrictive for it to exhaust the objects of perception: – sounds are public objects of audition, even if they are merely audible. They are audible bearers of perceptible attributes such as pitch, timbre, loudness, and duration. Multiple distinct sounds are audible at a time, sounds persist and survive change, and sounds can occlude and mask each other.

ONTOLOGICAL COMMITMENT identity judgments may even be perceptual, reifying esp. in the case of multi-modal perception: cf Matthew Nudds, Cross-modal object perception (A neo-Aristotelian approach?) and the papers by Casey O'Callaghan.

CONSTRUCTION OF THE PERCEPTUAL OBJECT (ie the thing I see as being so-and-so) In seeings-as, "the third argument of 'see as' attributes to a subject a non-conceptual way of seeing what the second argument refers to" (Mulligan 1999: 125): the "as" clause is a adverbial modification of the act of seeing. I defend this construal against the standard anti-adverbialist arguments. Individuating objects of seeing-as as manifestations, I argue, allows for both an ontologically plausible construal of the entities concerned, as for the 'right' logical relations among the ways they are seen. By distinguishing between the intentional and representational properties of seeing-as, we can explain why to undergo an emotion is indeed to stand in an intentional relation to value, but the relation is not belief nor does it involve any representation (thought) of value (Mulligan 2007: 209-210).

whether a adverbialist account of seeing-as can assuage some of the metaphysical worries people have raised about ephemeral or 'non-canonical' objects of perception such as flames, soap-bubbles, glimmers, highlights, reflections, echoes, shivers, atmospheric phenomena like rainbows and mirages, shadows, after-images, constellations, affordances and values.

> Many philosophers assume that reports employing 'see as' are made true by epistemic perceptions or, at the very least, by concept-involving perceptions. On the alternative view adopted here, the third argument of 'see as' attributest o a subject a non-conceptual way of seeing what the second argument refers to and all simple seeing involves seeing as. (Mulligan 1999: 125)

Seeing-as-the-same required for object constancy in the definition of visual acquaintance: "If x comes to be visually acquainted with y, then x sees y at t1 and then at t2 and sees y at t2 as the same object." (Mulligan 2007: 215) Also: x visually apprehends that p (has perceptual knowledge that p) if Sam sees that Maria is sad at t1 and then at t2 and identifies what he sees at t1 and at t2, where the identification is a mental act having the respective facts as its direct objects (Mulligan 2007: 216)

OBJECTS OF PERCEPTION An influential view among philosophers and psychologists of perception is that material objects are the intentional objects of perception1. This view is typically claimed to be in accordance with ordinary language (Strawson 1959), commonsensical intuitions, and phenomenology, and to be supported by various empirical data (Spelke 1993, Carey 2001, Bloom 1996 (Bloom 1996), Pylyshyn 1998 (Pylyshyn 1998)). This relative orthodoxy has been challenged on the grounds that perception of events (Mulligan, Simons, and Smith 1984), tropes (Campbell 1990, Kriegel 2004 (Kriegel 2004)), localized qualities (Clark 1993) (Clark 1993), and states

of affairs (Armstrong 1997) might in fact be more fundamental than the perception of material objects. OTHER SENSE MODALITIES: Cf. O?Callaghan (2011), Biggs, Matthen, and Stokes (2014) (Biggs et al. 2014).

general: (Matthen 2005) (on file)

5.3.4 The value of justification

5.3.5 Reactive epistemology

General motivation: the irreducibility of attitudes

general argument: - we can't get rid of attitudes anyway (speech act theory) - expressivism is inadequate as a theory of meaning; but there is no meaning anyway (Travis) - expressivism instead can teach us HOW our utterances acquire their meaning - Frege/Geach point: we need a logic of attitudes anyway (cf also deontic logic and credences) – check in particular: Hale (1986), Hale (1993) cf also discussion of disagreement in "perspectival disagreement"

cf 'intimation' and 'indication' in Husserl, Log Unter I (Simons 1995: 109): "What my statement to the neighbour *means* is not that *I believe* the day to be nice, but that the day itself is nice, whereas what my uttering intimates is that I believe the day to be nice."

The irreducibility of attitude in speech act theory

initial distinction between mode and content: Searle 1983, Tye 1997 psychological modes should be sharply distinguished from modes of presentation / 'aspectual shapes' (e.g. Searle 1983, Crane 2001 (Crane 2001), Crane 2013a (Crane 2013)). the distinction between psychological mode (or force, attitude) and content has been regularly acknowledged at least since the writings of Brentano (1874), Psychologie vom empirischen , Frege (1892) and Husserl (1901). It is intimately related to the contrast so regularly drawn from Frege onwards (e.g., Frege 1892, Stenius 1967 (Stenius 1967) and Dummett 1973, Frege) between the force and the content of speech acts such as asserting, ordering and questioning. reasons to believe in modes: (i) what Geach described as 'Frege's point' and used to discredit speech-act analyses of 'good' or 'intentional' (Geach 1960, 1965). For this reason, these observations continue to be at the centre of discussions surrounding the prospects of expressivism (Schroeder 2008). (ii) if we do not allow for independent variation of mode and content, questions, assertions, orders etc. (on the side of speech acts), as well as wonderings, beliefs, desires etc. (on that of mental states) could not have the same content; More specifically, it then becomes difficult to account for the logical and epistemological relations between mental states (e.g., Fodor 1979); (iii) the distinction is claimed to allow for a neat division of labour between semantics, which concerns itself with the nature of content and typically emphasizes its connections with truth, and pragmatics, whose subject-matter is the various 'uses' to which these contents are put (e.g., Pendlebury 1986). (Pendlebury 1986)

Performatives à la Austin: Austin says that they lack truth-values because they do not describe or report; but all he establishes is that they not JUST describe or report. But some think performatives have truth-values: Bach & Harnish (1992).

Hare against Searle on "I hereby promise that" (Searle 1969: 30–33) (file) (probably also in ,Searle (1962),): Hare (1970) (file) Warnock (1971)

WE NEED A LOGIC OF ATTITUDES ANYWAY (cf also deontic logic and credences) EMOTIONS AS MODES Edgington (1995) (file) claims that force indicators have their own compositionality, that "If he isn't at home, is he in college?" can be understood as having categorical content and conditional interrogative force

LOGIC OF ATTITUDES: BRANDOM, CF ALSO HIS TALK AT THE ST ANDREWS CONFERENCE

Consider (Parsons 2012: 52):

(**T**)   Attack at dawn!

Can you infer from (**T**) that someone commands something? In one sense, yes: (**T**) may be used to command, and in these cases "someone commands something" is true. In another sense, no: (**T**) no more entails "someone commands something" than "the weather is fine" entails "someone asserts something" (Parsons 2012: 53).

Validity and consistency – it's important to know which sets of instructions are consistent (Vranas 2010)

The Searlean conversion: from "attack at dawn!" to "I command you to attack at dawn". accepted by Lewis (1970: 220–226) and Lewis (1979: 234–236), Jackson & Pettit (1998: 248–249), Price (1988: ch3)

Attitudes determine what is said

Suppose we agree that normative sentences express attitudes. We then have to answer the question why some given sentence expresses the attitude with content *x* rather than an attitude with content *y*. The same-content theorists have an answer: because the sentence has content *x* (rather than *y*). The expressivists cannot say this, because in their view the sentence has content *x because* the attitude has the content *x*:

> …expressivists are committed to holding that the sentence 'grass is green and snow is white' gets its content *from* the belief it expresses. But in order for that to happen, it must *first* express a belief, and only *then* acquire a content. (Schroeder 2008a: 25)

Expressivists hold that evaluative sentences express, but do not report attitudes. But how do we know, of a given evaluative sentence, which attitude it expresses? To know this, we have to know the content of the attitude, as attitudes are individuated (at least partly) by their content. Two problems for expressivists:

- epistemic determination: in order to know the content of the attitude, we need to know the content of the sentence; but we cannot know this (e.g. by truth-conditions), *before* we know the content of the attitude;
- metaphysical determination: the sentence has the content it has because the attitude has the content *it* has, the sentence inherits its content from the attitude; hence, the content of the attitude cannot itself be determined by the content of the sentence.

Causal accounts are inadequate because of the liar problem: speakers must be able to express attitudes they don't in fact have.

Problem: This is a problem for everyone (also e.g. for Griceans) for whom the meaningfulness of sentences is not brute.

CHECK: how does Grice deal with liars?

Jackson & Pettit (1998) "…any explanation of how we English speakers came to use the voluntary sign 'good' for the attitude we do use it to express, according to expressivists, must allow that we *recognize* the attitude in question in us." (1998: 255) Because such a recognition will issue in a belief that we have the attitude in question, the predication of 'good' will be true iff our belief is true. They take the point to be quite general: for items having non-natural meaning, to express is to report the content of the attitude expressed (1998: 259) .

"…it makes no difference whether I say 'Shut the door' or I say 'I command that the door be shut' […]. But 'I command that the door be shut' obviously has truth conditions. And so it follows that 'Shut the door' has truth conditions too: it is true in *S*'s mouth at *t* if *S* did indeed command at *t* that the door be shut." (Jackson & Pettit 1998: 263)

van Roojen (1996): Gibbard cannot distinguish between an assertion of "I believe that *p*" expressing the belief that *p* and expressing the belief that I believe that *p*.

WHY ISN'T THIS EQUALLY AN OBJECTION TO GRICE?

Replies: Gibbard, Thinking How to Live, Schroeder (2008b)

Solution: assertability semantics: The sentence gets its content autonomously, in virtue of being associated with a proposition that the speaker is in a certain mental state. Which proposition this is (i.e. which belief the speaker required to have) is determined by its assertability conditions. These conditions require the speaker to have the

attitude in question, but do not state *that* the speaker has this attitude. These assertability conditions are different from the truth-conditions, because not every mistake needs to be linguistic:

> On the picture to which expressivists are committed [...] [, t]he primary job of the semantics is to assign to each atomic sentence a mental state – the state that you have to be in, in order for it to be permissible for you to assert that sentence. (Schroeder 2008a: 33)

Problem: Analogy with fear – isn't there an analoguous distinction to be made in the belief / disapproval case?: My fear of this spider is correct (i) iff the spider is dangerous; (ii) iff I believe the spider to be dangerous. Why should only the latter give the content?

Assertoric force

valid:

A1 You will be happy if this damn Kaplan comes.
A2 This damn Kaplan comes.
A3 You will be happy.

invalid:

B1 The damn Kaplan comes if he wakes up on time.
B2 He wakes up on time.
B3 The damn Kaplan comes.

ASSERTORIC FORCE A FORTIORI inferences, begging the question in a strict sense: where the conclusion is already asserted in the premisses (true for conjunction elimination, not true for MP).

Attitudes are what we report

The basic puzzle: a sentence such like "the weather is fine" may be used not only to talk about the weather, but also to talk about us – it may successfully be deployed to characterise Sam as believing, desiring, wishing, Maria as hoping, expecting or remembering and Fred as saying, stating, judging, supposing or asking. This use of sentences about the weather in the classification and characterisation of mental states raises a problem: what is the relation between these states and the state of the weather called upon to characterise them? In virtue of what are these characterisations true?

Many people have thought that a useful first step towards answering these questions is to focus on just one type of mental state, belief, and one type of characterisation, the stating of truth-conditions. This approach has a number of drawbacks:

- Starting with belief is insufficiently general: beliefs are just one kind of mental states, and the conceptual apparatus that is useful to analyse them may be too parochial to be of any help in the analysis of other mental states.
- Taking beliefs as paradigm mental states may tempt us to reify truth-conditions into self-standing objects, propositions or contents. These propositions, introduced as objects of beliefs, are too fine-grained to be plausibly taken to be the objects of the other 'propositional' attitudes.
- Neglecting other mental states than beliefs, we misconstrue the phenomenal, normative and affective properties of the latter: they come to be seen as phenomenally, normatively and affectively inert, and we end up with a meagre and lifeless conception of our desire for, and love of, truth.
- Focussing too much on belief, we make it hard to understand how belief connects to other mental states. Especially mental states that are entanglements of belief with other states fall out of the picture.

Rather than starting with beliefs, we should first develop a general theory of correctness conditions. The fundamental notion, I submit, is representation-as.

CHECK: consistency requirements / conditional excluded middle in deontic logic

CHECK: consistency requirements in value theory

CHECK: consistency requirements in Broome

Bibliography

Armstrong, David M., 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

Atkin, Albert, 2006. There's no place like "here" and no time like "now" . *American Philosophical Quarterly* 43(3): 271–280.

Bach, Kent & Robert M. Harnish, 1992. How Performatives Really Work: A Reply to Searle (1989). *Linguistics and Philosophy* 15(1): 93–110.

Barwise, Jon K., 1988. On the Model Theory of Common Knowledge. Technical Report, Center for the Study of Language and Information, Stanford, California. Reprinted in Barwise (1989: 201–220).

Barwise, Jon K. (editor) , 1989. *The Situation in Logic*. Stanford, California: CSLI Publications.

Barwise, Jon K. & John R. Perry (editors) , 1983. *Situations and Attitudes*. 1999 (2nd) edition. Cambridge, Massachusetts: The MIT Press. Reissued 1999 in the *David Hume Series of Philosophy and Cognitive Science Reissues* series of CSLI Press, Stanford.

Bäuerle, Rainer, Urs Egli & Arnim von Stechow (editors) , 1979. *Semantics from Different Points of View*. Number 6 in Springer Series in Language and Communication, Berlin: Springer Verlag.

Biggs, Stephen, 2017. Modal Conventionalism and Textbook Analyticities. In Fischer & Leon (2017b), pp. 155–178.

Biggs, Stephen, Mohan Matthen & Dustin R. Stokes, 2014. Sorting the Senses. In Stokes, Dustin R., Mohan Matthen & Stephen Biggs (editors) *Perception and its Modalities*, pp. 1–21. Oxford: Oxford University Press.

Blackburn, Simon, 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press.

Bloom, Paul, 1996. Possible Individuals in Language and Cognition. *Current Directions in Psychological Science* 5: 90–94.

Blum, Philipp, 2011. Racisme et bêtise. In Tappolet, Christine, Fabrice Teroni & Anita Konzelmann Ziv (editors) *Les Ombres de l'Âme. Penser les émotions négatives*, pp. 145–159. Genève: Éditions Markus Haller. Published under the name of "Keller" .

Boghossian, Paul Artin & Christopher Peacocke (editors) , 2000. *New Essays on the A Priori*. Oxford: Clarendon Press.

BonJour, Laurence, 2009. *Epistemology. Classic Problems and Contemporary Responses*. 2 edition. Lanham, Maryland: Rowman & Littlefield Publishers, Inc. First edition: **?**.

Brady, Michael Sean, 2010. Virtue, emotion, and attention. *Metaphilosophy* 41(1–2): 115–131.

Brentano, Franz, 1874. *Psychologie vom Empirischen Standpunkte*, volume I. Leipzig: Dunker und Humblot. Second edition: Brentano (1924).

Brentano, Franz, 1924. *Psychologie vom Empirischen Standpunkt*, volume I. 2 edition. Leipzig: Felix Meiner Verlag. Edited by Oskar Kraus.

Broad, Charlie Dunbar, 1923. *Scientific Thought*. International Library of Psychology, Philosophy and Scientific Method, London: Kegan Paul, Trench, Trübner & Co.

Broad, Charlie Dunbar, 1925. *The Mind and Its Place In Nature*. International Library of Psychology, Philosophy and Scientific Method, London: Kegan Paul, Trench, Trübner & Co.

Brueckner, Anthony, 1986. Brains in a Vat. *The Journal of Philosophy* 83(3): 148–167. Reprinted in Brueckner & Ebbs (2012: 9–27).

Brueckner, Anthony, 2005. Fallibilism, Underdetermination, and Skepticism. *Philosophy and Phenomenological Research* 71(2): 384–391.

Brueckner, Anthony & Gary Ebbs (editors) , 2012. *Debating Self-Knowledge*. Cambridge: Cambridge University Press.

Bueno, Otávio, 2017. The Epistemology of Modality and the Epistemology of Mathematics. In Fischer & Leon (2017b), pp. 67–84.

Burge, Tyler, 1988. Individualism and Self-Knowledge. *The Journal of Philosophy* 85(11): 649–663. Reprinted in Burge (2013: 55–67).

Burge, Tyler, 1993. Content Preservation. *The Philosophical Review* 102(4): 457–458. Reprinted in Villanueva (1995: 271–300) and in Burge (2013: 229–253).

Burge, Tyler, 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Burge, Tyler, 2013. *Cognition Through Understanding*. Number 3 in Philosophical Essays, Oxford: Clarendon Press.

Burnyeat, Myles F. (editor) , 1983. *The Skeptical Tradition*. Berkeley, California: University of California Press.

Burri, Alex, 1994. *Hilary Putnam*. Frankfurt a.M.: Campus Verlag.

Byrne, Alex & Heather Logue (editors) , 2009. *Disjunctivism: Contemporary Readings*. Cambridge, Massachusetts: The MIT Press.

Carriero, John P., 1987. The First Meditation. *Pacific Philosophical Quarterly* 68: 222–248. Cited after reprint in Chappell (1997: 1–31).

Castañeda, Hector-Neri, 1959. The Logic of Obligation. *Philosophical Studies* 10(2): 17–23.

Castañeda, Hector-Neri, 1964. Correction to "The Logic of Obligation" [Castañeda (1959)] (A Reply). *Philosophical Studies* 15(1–2): 25–28.

Chappell, Vere C. (editor) , 1997. *Descartes's Meditations, Critical Essays*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.

Chisholm, Roderick M., 1952. Intentionality and the Theory of Signs. *Philosophical Studies* 3(4): 56–63.

Chisholm, Roderick M., 1977. *Theory of Knowledge*. 2 edition. Foundations of Philosophy Series, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Clark, Austen, 1993. *Sensory Qualities*. Oxford: Clarendon Press.

Cohen, Stewart, 1988. How to Be a Fallibilist. In Tomberlin, James E. (editor) *Philosophical Perspectives 2: Epistemology*, pp. 91–123. Oxford: Basil Blackwell Publishers.

Cook Wilson, John, 1926. *Statement and Inference. With other Philosophical Papers*. Oxford: Clarendon Press. Edited from the mss. by A.S.L. Farquharson.

Crane, Tim, 2001. *Elements of Mind, An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.

Crane, Tim, 2005. The Problem of Perception. In Zalta, Edward N. (editor) *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information. Version of March 8, 2005.

Crane, Tim, 2013. *The Objects of Thought*. New York: Oxford University Press.

Currie, Gregory, 2002. Desire in Imagination. In Gendler & Hawthorne (2002), pp. 201–222.

Davidson, Donald & Gilbert H. Harman (editors) , 1972. *Semantics of Natural Language*. Number 40 in Synthese Library, Dordrecht: D. Reidel Publishing Co.

Davies, David, 1995. Putnam's Brain-Teaser. *Canadian Journal of Philosophy* 25: 203–228.

Davis, J.W., D.T. Hockney & W. Kent Wilson (editors) , 1969. *Philosophical Logic*. Number 20 in Synthese Library, Dordrecht: D. Reidel Publishing Co.

Dell'Utri, Massimo, 1990. Choosing Conceptions of Realism: The Case of the Brains in a Vat. *Mind* 99: 79–90.

Deonna, Julien Amos & Fabrice Teroni, 2008. *Qu'est-ce qu'une émotion?* Chemins Philosophiques, Paris: Librairie philosophique Jean Vrin.

Deonna, Julien Amos & Fabrice Teroni, 2012. *The Emotions. A Philosophical Introduction*. London: Routledge. Substantially enlarged edition and translation of Deonna & Teroni (2008).

Deonna, Julien Amos & Fabrice Teroni, 2015. Emotions as Attitudes. *Dialectica* 69(3): 293–311. Special issue "Beyond Perceptualism", edited by Sabine A. Döring and Anika Lutz.

Devlin, Keith J., 1991. *Logic and Information*. Cambridge: Cambridge University Press.

Dretske, Fred I., 1970. Epistemic Operators. *The Journal of Philosophy* 67(24): 1007–1023. Reprinted in Dretske (2000: 30–47).

Dretske, Fred I., 1971. Conclusive Reasons. *Australasian Journal of Philosophy* 49: 1–22. Reprinted in Dretske (2000: 3–29).

Dretske, Fred I., 1981. *Knowledge and the Flow of Information*. Cambridge, Massachusetts: The MIT Press. Reissued 1999 in the *David Hume Series of Philosophy and Cognitive Science Reissues* series of CSLI Press, Stanford.

Dretske, Fred I., 2000. *Perception, Knowledge and Belief: Selected Essays*. Cambridge: Cambridge University Press.

Duc, Ho Ngoc, 2001. *Resource-bounded reasoning about knowledge*. PhD dissertation, University of Leipzig, Faculty of Mathematics and Informatics, Leipzig.

Dummett, Michael A.E., 1973. *Frege: Philosophy of Language*. London: Duckworth. 1981 (2nd) edition with Harvard University Press, Cambridge MA.

Dummett, Michael A.E., 2002. 'Yes', 'No' and 'Can't Say'. *Mind* 111(442): 289–295.

Edgington, Dorothy, 1995. On Conditionals. *Mind* 104(413): 235–329.

Empiricus, Sextus, 2000. *Outlines of Scepticism*. 2 edition. Cambridge: Cambridge University Press. Edited by Julia Annas and Jonathan Barnes.

Field, Hartry, 1994. Disquotational Truth and Factually Defective Discourse. *The Philosophical Review* 103: 405–452. Reprinted in Field (2001: 222–258).

Field, Hartry, 2000. Apriority as an Evaluative Notion. In Boghossian & Peacocke (2000), pp. 117–149. Reprinted in Field (2001: 361–387).

Field, Hartry, 2001. *Truth and the Absence of Facts*. Oxford: Clarendon Press.

Fischer, Bob, 2017. Modal Empiricism: Objection, Reply, Proposal. In Fischer & Leon (2017b), pp. 273–280.

Fischer, Bob & Felipe Leon, 2017a. Introduction. In Fischer & Leon (2017b), pp. 1–6.

Fischer, Bob & Felipe Leon (editors) , 2017b. *Modal Epistemology After Rationalism*. Number 378 in Synthese Library, Cham: Springer Verlag.

Fodor, Jerry A., 1984. Semantics, Wisconsin Style. *Synthèse* 59(3): 231–250. Reprinted in Silvers (1989) and in Fodor (1994: 31–50).

Fodor, Jerry A., 1994. *A Theory of Content and Other Essays*. Cambridge, Massachusetts: The MIT Press.

Francescotti, Robert M. (editor) , 2014. *A Companion to Intrinsic Properties*. Berlin: Walter de Gruyter.

Frege, Gottlob, 1879–1891. Logik. In Frege (1969), pp. 1–8. Translated in Frege (1979: 1–8).

Frege, Gottlob, 1969. *Nachgelassene Schriften*. Hamburg: Felix Meiner Verlag. Edited by Hermes, Hans and Kambartel, Friedrich and Kaulbach, Friedrich; 2nd edition: Frege (1983).

Frege, Gottlob, 1979. *Posthumous Writings*. Oxford: Basil Blackwell Publishers. Edited by Hermes, Hans and Kambartel, Friedrich and Kaulbach, Friedrich; translation of (Frege 1969) by Peter Long and Roger Whit.

Frege, Gottlob, 1983. *Nachgelassene Schriften*. 2 edition. Hamburg: Felix Meiner Verlag. Edited by Hermes, Hans and Kambartel, Friedrich and Kaulbach, Friedrich.

French, Peter A., Theodore E. Uehling, Jr. & Howard K. Wettstein (editors) , 1977. *Midwest Studies in Philosophy 2: Contemporary Perspectives in the Philosophy of Language*. Minneapolis, Minnesota: University of Minnesota Press. Second, enlarged edition: French et al. (1979).

French, Peter A., Theodore E. Uehling, Jr. & Howard K. Wettstein (editors) , 1979. *Midwest Studies in Philosophy 2: Contemporary Perspectives in the Philosophy of Language*. 2 edition. Minneapolis, Minnesota: University of Minnesota Press. First edition: French et al. (1977).

Gendler, Tamar Szabó, 2000. The Puzzle of Imaginative Resistance. *The Journal of Philosophy* 97(2): 55–81. Reprinted in Gendler (2010: 155–178).

Gendler, Tamar Szabó, 2010. *Intuition, Imagination, and Philosophical Methodology*. Oxford: Oxford University Press.

Gendler, Tamar Szabó & John Hawthorne (editors) , 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.

Gettier, Edmund L. III, 1963. Is Justified True Belief Knowledge? *Analysis* 23: 121–123.

Gibbard, Allan F., 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Cambridge, Massachusetts: Harvard University Press.

Gibbard, Peter J., 2002. Price and Rumfitt on Rejective Negation and Classical Logic. *Mind* 111(442): 297–303.

Gochet, Paul & Pascal Gribomont, 2006. Epistemic Logic. In *Handbook of the History of Logic. Volume 7: Logic and the Modalities in the 20th Century*, pp. 99–195. Amsterdam: North-Holland Publishing Co.

Goldie, Peter, 2000. *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.

Goldman, Alvin I., 1967. A Causal Theory of Knowing. *The Journal of Philosophy* 64: 357–372.

Gundersen, Lars Bo, 2010. Tracking, Epistemic Dispositions and the Conditional Analysis. *Erkenntnis* 72(3): 353–364.

Hale, Bob, 1986. The Compleat Projectivist [critical notice of Blackburn (1984)]. *The Philosophical Quarterly* 36(142): 65–84.

Hale, Bob, 1993. Can there be a Logic of Attitudes? In Haldane, John & Crispin Wright (editors) *Reality, Representation & Projection*, pp. 337–364. Mind Association Occasional Series, Oxford: Oxford University Press.

Halpern, Joseph Y., 1996. Should Knowledge Entail Belief? *The Journal of Philosophical Logic* 25(5): 483–494.

Hanrahan, Rebecca Roman, 2017. Imagination, Possibility, and Powers. In Fischer & Leon (2017b), pp. 197–220.

Hare, Richard M., 1970. Meaning and Speech Acts. *The Philosophical Review* 79(1): 3–24.

Hawke, Peter, 2017. Can Modal Skepticism Defeat Humean Skepticism? In Fischer & Leon (2017b), pp. 281–308.

Hawthorne, John, 2004. *Knowledge and Lotteries*. Oxford: Clarendon Press.

Hawthorne, John, 2005. Vagueness and the Mind of God. *Philosophical Studies* 122(1): 1–25. Reprinted in Hawthorne (2006: 165–183).

Hawthorne, John, 2006. *Metaphysical Essays*. Oxford: Clarendon Press.

Hintikka, Jaakko, 1962. *Knowledge and Belief*. Ithaca, New York: Cornell University Press.

Hintikka, Jaakko, 1968. Epistemic Logic and the Methods of Philosophical Analysis. *Australasian Journal of Philosophy* 46: 37–51. Reprinted in Hintikka (1969: 3–21).

Hintikka, Jaakko, 1969. *Models for Modalities: Selected Essays*. Number 23 in Synthese Library, Dordrecht: D. Reidel Publishing Co.

Hintikka, Jaakko, 1970. On Semantic Information. In Yourgrau, Wolfgang & Allen D. Breck (editors) *Physics, Logic, and History. Based on the First International Colloquium held at the University of Denver, May 16-20, 1966*, pp. 147–168. New York: Plenum Press. Reprinted in Hintikka & Suppes (1970: 3–27).

Hintikka, Jaakko & Patrick Suppes (editors) , 1970. *Information and Inference*. Number 28 in Synthese Library, Dordrecht: D. Reidel Publishing Co.

Hochberg, Herbert, 1988. A Refutation of Moderate Nominalism. *Australasian Journal of Philosophy* 66(2): 188–207. Elaborated into Hochberg (2001).

Hochberg, Herbert, 2001. A Refutation of Moderate Nominalism. In *Russell, Moore and Wittgenstein. The Revival of Realism*, pp. 175–204. Number 1 in Philosophische Analyse / Philosophical Analysis, Heusenstamm b. Frankfurt: Ontos Verlag.

Hoffman, Martin L., 2000. *Empathy and Moral Development*. Cambridge: Cambridge University Press.

Holt, Edwin Bissell, 1912. The Place of Illusory Experience in a Realistic World. In Holt, Edwin Bissell, Walter Taylor Marvin, William Pepperell Montague, Ralph Barton Perry, Walter Boughton. Pitkin & Edward Gleason Spaulding (editors) *The new realism: Cooperative studies in philosophy*, pp. 303–373. New York: MacMillan Publishing Co.

Hossack, Keith G., 2007. *The Metaphysics of Knowledge*. Oxford: Oxford University Press.

Humberstone, I. Lloyd, 2000. The Revival of Rejective Negation. *The Journal of Philosophical Logic* 29(4): 331–381.

Jackson, Frank, 1977. *Perception: a Representative Theory*. Cambridge: Cambridge University Press.

Jackson, Frank, Philip Pettit & Michael A. Smith (editors) , 2004. *Mind, Morality, and Explanation. Selected Collaborations*. Oxford: Clarendon Press.

Jackson, Frank & Phillip Pettit, 1998. A Problem for Expressivism. *Analysis* 58: 239–251. Reprinted in Jackson et al. (2004: 252–265).

Johnston, Mark, 2004. The Obscure Object of Hallucination. *Philosophical Studies* 120(1–3): 113–183. Reprinted in Byrne & Logue (2009: 207–270).

Kenny, Anthony John Patrick, 1963. *Action, Emotion and the Will*. London: Routledge & Kegan Paul. 2nd edition: Kenny (2003).

Kenny, Anthony John Patrick, 2003. *Action, Emotion and the Will*. 2 edition. London: Routledge & Kegan Paul. First edition: Kenny (1963).

Kriegel, Uriah, 2004. Trope Theory and the Metaphysics of Appearances. *American Philosophical Quarterly* 41(1): 5–20.

Kripke, Saul A., 1980. *Naming and Necessity*. Oxford: Basil Blackwell Publishers.

Krödel, Thomas, 2017. Modal Knowledge, Evolution, and Counterfactuals. In Fischer & Leon (2017b), pp. 179—196.

Kung, Peter, 2017. Personal Identity Without Too Much Science Fiction. In Fischer & Leon (2017b), pp. 133–154.

Lehrer, Keith, 1975. *Analysis and Metaphysics. Essays in Honor of R.M. Chisholm*. Number 4 in Philosophical Studies Series, Dordrecht: D. Reidel Publishing Co. Second edition: Lehrer (1994).

Lehrer, Keith, 1994. *Analysis and Metaphysics*. 2 edition. Dordrecht: D. Reidel Publishing Co. First edition: Lehrer (1975).

Lemmon, Edward John, 1965. *Beginning Logic*. London: Thomas Nelson and Sons Ltd.

Lenzen, Wolfgang, 1978. *Recent Work in Epistemic Logic*. Number 30, 1 in Acta Philosophica Fennica, Helsinki: Societas Philosophica Fennica, Akateeminen Kirjakauppa.

Lenzen, Wolfgang, 1979. *Glauben, Wissen und Wahrscheinlichkeit*. Wien: Springer Verlag.

Leon, Felipe, 2017. From Modal Skepticism to Modal Empiricism. In Fischer & Leon (2017b), pp. 247–262.

Lewis, David, 1970. General Semantics. *Synthèse* 22(1–2): 18–67. Republished in Davidson & Harman (1972: 169–218), in Partee (1976: 1–50) and, with a postscript (Lewis 1983b), in Lewis (1983a: 189–230).

Lewis, David, 1979. Scorekeeping in a Language Game. *The Journal of Philosophical Logic* 8(3): 339–359. Republished in Bäuerle et al. (1979: 172–187) and in Lewis (1983a: 233-249).

Lewis, David, 1983a. *Philosophical Papers*, volume 1. Oxford: Oxford University Press.

Lewis, David, 1983b. Postscript to Lewis (1970). In Lewis (1983a), pp. 230–232.

Lewis, David, 1996. Elusive Knowledge. *Australasian Journal of Philosophy* 74: 549–567. Reprinted in Lewis (1999: 418–445).

Lewis, David, 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David & Rae Langton, 1998. Defining "intrinsic" . *Philosophy and Phenomenological Research* 58(2): 333–345. Reprinted in Lewis (1999: 116–132) and in Francescotti (2014: 17–30).

Martin, M.G.F., 2010. What's in a Look? In Nanay, Bence (editor) *Perceiving the World*, pp. 160–225. Oxford: Oxford University Press.

Martin, M.G.F., 2015. Moore's Dilemma. In Coates, Paul & Sam Coleman (editors) *Phenomenal Qualities. Sense, Perception, and Consciousness*, pp. 147–180. Oxford: Oxford University Press.

Matthen, Mohan, 2005. *Seeing, Doing, and Knowing A Philosophical Theory of Sense Perception*. Oxford: Clarendon Press.

McGee, Vann, 1985. A Counterexample to Modus Ponens. *The Journal of Philosophy* 82(9): 462–471.

Moore, George Edward, 1912. *Ethics*. Oxford: Oxford University Press.

Moore, George Edward, 1914. The Status of Sense-Data. *Proceedings of the Aristotelian Society* 14: 355–380. Reprinted in Moore (1922: 168–196).

Moore, George Edward, 1922. *Philosophical Studies*. London: Routledge & Kegan Paul.

Moran, Richard, 1994. The Expression of Feeling in Imagination. *The Philosophical Review* 103(1): 75–106.

Morris, Charles W., 1939. *Foundations of the Theory of Signs*. Number I, 2 in International Encyclopedia of Unified Science, Chicago, Illinois: University of Chicago Press. Reprinted in Neurath et al. (1955: 77–136).

Müller, Olaf L., 2001. Does Putnam's Argument Beg the Question Against the Skeptics? Bad News for Radical Scepticism. *Erkenntnis* 54(3): 299–320.

Mulligan, Kevin, 1998. From Appropriate Emotions to Values. *The Monist* 81(1): 161–188.

Mulligan, Kevin, 1999. Justification, Rule-Breaking and the Mind. *Proceedings of the Aristotelian Society* 99: 123–139.

Mulligan, Kevin, 2006a. Ascent, Propositions, and Other Formal Objects. *Grazer Philosophische Studien* 72: 29–48. "Propositions", ed. by Massimiliano Carrara and Elisabetta Sacchi.

Mulligan, Kevin, 2006b. Facts, Formal Objects and Ontology. In Bottani, Andrea & Richard Davies (editors) *Modes of Existence. Papers in Ontology and Philosophical Logic*, pp. 31–46. Number 5 in Philosophische Forschung / Philosophical Research, Heusenstamm b. Frankfurt: Ontos Verlag. Papers from the conference "On what (perhaps) there is", held May 2005 at the University of Bergamo.

Mulligan, Kevin, 2007. Intentionality, Knowledge and Formal Objects. In Rønnow-Rasmussen, Toni, Björn Petersson, Jonas Josefsson & Dan Egonsson (editors) *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Lund: Philosophy Department. URL http://www.fil.lu.se/hommageawlodek.

Neurath, Otto, Rudolf Carnap & Charles W. Morris (editors) , 1955. *International Encyclopedia of Unified Science*, volume I, 1–5. Chicago, Illinois: University of Chicago Press.

Nolan, Daniel Patrick, 2017. Are There Any Worldly States of Affairs? Yes. In Barnes, Elizabeth (editor) *Current Controversies in Metaphysics*, pp. 81–91. New York: Routledge.

Nozick, Robert, 1981. *Philosophical Explanations*. Cambridge, Massachusetts: Harvard University Press.

Nunn, T.Percy, 1910. Are Secondary Qualities Independent of Perception? *Proceedings of the Aristotelian Society* 10: 191–218.

O'Callaghan, Casey, 2016. Objects for Multisensory Perception. *Philosophical Studies* 173(5): 375–405.

O'Shaughnessy, Brian, 2000. *Consciousness and the World*. Oxford: Oxford University Press.

Parsons, Josh, 2012. Cognitivism about imperatives. *Analysis* 72(1): 49–54.

Partee, Barbara Hall (editor) , 1976. *Montague Grammar*. New York: Academic Press.

Pendlebury, Michael J., 1986. Against the Power of Force: Reflections on the Meaning of Mood. *MInd* 95: 361–373.

Pérez-Otero, Manuel, 2011. Modest Skepticism and Question Begging Proper. *Grazer Philosophische Studien* 83: 9–32.

Poser, Hans, 2003. *René Descartes: Eine Einführung*. Stuttgart: Philipp Reclam jun. GmbH.

Price, Huw, 1983. Sense, Assertion, Dummett and Denial. *Mind* 92: 161–173.

Price, Huw, 1988. *Facts and the Function of Truth*. Oxford: Basil Blackwell Publishers.

Pryor, James, 2000. The Skeptic and the Dogmatist. *Noûs* 34: 517–549.

Putnam, Hilary, 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

Pylyshyn, Zenon W., 1998. Situation Vision in the World. *Trends in Cognitive Science* 2: 10–18.

Rasmussen, Joshua, 2017. Extending Modal Vision. In Fischer & Leon (2017b), pp. 115–132.

Récanati, François, 1987. Literalness and Other Pragmatic Principles. *Behavioral and Brain Sciences* 10: 729–730.

Roca-Royes, Sonia, 2017. Similarity and Possibility: An Epistemology of *de re* Possibility for Concrete Entities. In Fischer & Leon (2017b), pp. 221–246.

Rollins, C.D. (editor) , 1964. *Knowledge and Experience: Proceedings of the 1962 Oberlin Colloquium in Philosophy*. Pittsburgh, Pennsylvania: University of Pittsburgh Press.

van Roojen, Mark, 1996. Expressivism and Irrationality. *The Philosophical Review* 105(3): 322–335.

Ross, James Francis, 1975. Testimonial Evidence. In Lehrer (1975), pp. 35–56. Second edition: Lehrer (1994).

Rumfitt, Ian, 2000. 'Yes' and 'No'. *Mind* 109(436): 781–823.

Rumfitt, Ian, 2002. Unilateralism Disarmed: A Reply to Dummett (2002) and Gibbard (2002). *Mind* 111(442): 305–321.

Schroeder, Mark, 2008a. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Clarendon Press.

Schroeder, Mark, 2008b. Expression for Expressivists. *Philosophy and Phenomenological Research* 76(1): 86–116. Reprinted in Schroeder (2015: 31–54).

Schroeder, Mark, 2015. *Expressing Our Attitudes. Explanation and Expression in Ethics, Volume 2*. Oxford: Oxford University Press.

Searle, John R., 1962. Meaning and Speech Acts. *The Philosophical Review* 71(4): 423–432. Reprinted in Rollins (1964: 28–37).

Searle, John R., 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, John R., 1989. How Performatives Work. *Linguistics and Philosophy* 12(5): 535–558.

Searle, John R., 1992. *The Rediscovery of the Mind*. Cambridge, Massachusetts: The MIT Press.

Sellars, Wilfrid, 1956. The Myth of the Given: Three Lectures on Empiricism and the Philosophy of Mind. In Feigl, Herbert & Michael Scriven (editors) *Minnesota Studies in the Philosophy of Science, Volume I: Foundations of Science & the Concepts of Psychology and Psychoanalysis*, pp. 253–329. Minneapolis, Minnesota: University of Minnesota Press. Reprinted, under the tile "Empiricism and the Philosophy of Mind", in Sellars (1963: 127–196) and in book form as Sellars (1997).

Sellars, Wilfrid, 1963. *Science, Perception and Reality*. London: Routledge & Kegan Paul. Reprinted as Sellars (1991).

Sellars, Wilfrid, 1991. *Science, Perception and Reality*. Atascadero, California: Ridgeview Publishing Co. Reprint of Sellars (1963).

Sellars, Wilfrid, 1997. *Empiricism and the Philosophy of Mind*. Cambridge, Massachusetts: Harvard University Press. Republication of Sellars (1956), with an introduction by Richard Rorty and a study guide by Robert Brandom.

Shalkowski, Scott A., 2017. Modal Epistemology Without Detours. In Fischer & Leon (2017b), pp. 47–66.

Silvers, Stuart (editor) , 1989. *Rerepresentation: Readings in the Philosophy of Language*. Number 40 in Philosophical Studies Series, Dordrecht: D. Reidel Publishing Co.

Simons, Peter M., 1995. Meaning and Language. In Smith, Barry & David Woodruff Smith (editors) *The Cambridge Companion to Husserl*, pp. 106–137. Cambridge Companions to Philosophy, Cambridge: Cambridge University Press.

Stalnaker, Robert C., 1999. The Problem of Logical Omniscience, II. In *Context and Content: Essays on Intensionality Speech and Thought*, pp. 255–273. Oxford: Oxford University Press.

Stampe, Dennis W., 1977. Toward a Causal Theory of Linguistic Representation. In French et al. (1977), pp. 42–63. Reprinted in French et al. (1979: 81–102).

Stanley, Jason, 2005. Fallibilism and concessive knowledge attributions. *Analysis* 65(2): 126–131.

Stenius, Erik, 1967. Mood and Language-Game. *Synthèse* 17(1): 254–274. Reprinted in Davis et al. (1969: 251–271).

Stine, Gail Caldwell, 1976. Skepticism, Relevant Alternatives, and Deductive Closure. *Philosophical Studies* 29: 249–260.

Stroud, Barry, 1984. *The Significance of Philosophical Scepticism*. Oxford: Oxford University Press.

Tahko, Tuomas E., 2017. Empirically-Informed Modal Ratioanlism. In Fischer & Leon (2017b), pp. 29–46.

Teroni, Fabrice, 2007. Emotions and Formal Objects. *Dialectica* 61(3): 395–415.

Vaidya, Anand Jayprakash, 2017. Modal Knowledge: Beyond Rationalism and Empiricism. In Fischer & Leon (2017b), pp. 85–114.

Villanueva, Enrique (editor) , 1995. *Philosophical Issues 6: Content*. Atascadero, California: Ridgeview Publishing Co.

Vranas, Peter B.M., 2010. In defence of imperative inference. *The Journal of Philosophical Logic* 39(1): 59–71.

Walton, Kendall L., 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Massachusetts: Harvard University Press.

Walton, Kendall L., 1994. Morals in Fiction and Fictional Morality. *Proceedings of the Aristotelian Society, Supplementary Volume* 68: 27–50.

Warnock, Geoffrey J., 1971. Hare on Meaning and Speech Acts [on Hare (1970)]. *Proceedings of the Aristotelian Society* 72: 80–84.

Weatherson, Brian, 2004. Morality, Fiction and Possibility. *Philosophers' Imprint* 4(3).

Williams, Bernard Arthur Owen, 1978. *Descartes − The Project of Pure Enquiry*. London: Penguin Books.

Williamson, Timothy, 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, Timothy, 2002. Necessary Existents. In O'Hear, Anthony (editor) *Logic, Thought and Language*, pp. 233–251. Number 51 in Royal Institute of Philosophy Supplement, Cambridge: Cambridge University Press.

Wright, Crispin, 2004. Warrant for Nothing (and Foundations for Free)? *Proceedings of the Aristotelian Society, Supplementary Volume* 78: 167–212.

Yablo, Stephen, 2002. Coulda, Woulda, Shoulda. In Gendler & Hawthorne (2002), pp. 441–492.