Contrastive Self-Knowledge and the McKinsey Paradox[1]

Sarah Sawyer

Final Draft. For the official version please refer to:

S. Goldberg (ed.) *Externalism and Skepticism* (Cambridge: Cambridge University Press, 2015)

1. Introduction

According to the traditional account of knowledge, knowledge is a two-place, categorical relation between a subject and a proposition known: Ksp (S knows that p). According to a contrastive account of knowledge, knowledge is a three-place, contrastive relation holding between a subject, a proposition and a contrast class: Kspq (S knows that p rather than that q). There is reason to believe that knowledge is contrastive.[2] Self-knowledge is a kind of knowledge. This means that there is reason to believe that self-knowledge is contrastive.[3] In this paper I will argue that understanding self-knowledge as contrastive provides the means to overcome a perceived tension between two otherwise plausible theses: the thesis that we have privileged access to the contents of our own psychological states; and the thesis that our psychological states are anti-individualistically individuated—that they are individuated with essential reference to our environment. The apparent tension arises, in brief, because privileged access to the contents of one's anti-individualistically individuated psychological states appears to provide broadly a priori knowledge of the environmental conditions which serve in part to individuate those states. This strikes most as wildly implausible—how could we have broadly a priori knowledge of our environment? It also seems to provide an implausibly quick response to the external world sceptic. This is the core of the so-called 'McKinsey Paradox'. I will argue that a contrastive account of self-knowledge shows how we

[1] This paper was written in connection with Project FFI2012-38908-C02-02, 'Self-Knowledge, Expression and Transparency' funded by MINECO, a branch of the Spanish Government.

[2] See for example Karjalainen & Morton (2003), Schaffer (2004, 2005, 2007, and 2008), Morton & Karjalainen (2008), and Schaffer & Knobe (2012).

[3] See Sawyer (2014) for a defence of the view.

can have privileged access to our anti-individualistically individuated psychological states without the unwanted implication that we thereby have broadly a priori knowledge of our environment. Consequently, the perceived tension between privileged access and anti-individualism disappears. Moreover, a contrastive account of self-knowledge clarifies the extent to which we can have privileged access to our anti-individualistically individuated psychological states, and shows why such privileged access cannot provide an easy answer to the external world sceptic.

The structure of the paper is as follows. In section 2, I provide a brief characterization of the thesis of anti-individualism. In section 3, I provide a brief characterization of the thesis of privileged access. In section 4, I explain the apparent tension between the two theses as illustrated by the McKinsey paradox. In section 5, I set the stage for a contrastive solution to the McKinsey paradox and introduce a contrastive account of self-knowledge. In section 6, I show how a contrastive account of self-knowledge can solve the McKinsey paradox. This then provides an understanding of how—and the extent to which—we can have privileged access to our anti-individualistically individuated psychological states, and shows why such privileged access cannot be used to refute external world scepticism.

2. Anti-individualism

Anti-individualism is the thesis that certain of an individual's non-representational, causal relations to objective properties in her environment are constitutive of her representational mental kinds, and hence of what she can represent in thought. Anti-individualism implies that two individuals who are intrinsic physical duplicates may well instantiate different representational mental kinds—may well have different thoughts—in virtue of being related to different objective properties in their respective environments. This contrasts with individualism, according to which intrinsic physical duplicates necessarily instantiate the very same representational mental kinds—necessarily have the same thoughts—because, according to individualism, the thoughts an individual has are constitutively what they are independently of any relations she has to her particular environment.[4]

[4] The dispute between individualists and anti-individualists turns on the question of whether relations to objective properties in one's environment play a constitutive role in determining one's representational mental kinds. It is consistent with both positions that relations to such properties play a causal role.

Anti-individualism is notoriously supported by Twin Earth thought experiments.[5] These invite reflection on counterfactual scenarios in which an individual's intrinsic physical make-up is hypothesized to remain constant while the wider environment in which she is embedded is hypothesized to differ. According to the anti-individualist, the hypothesized differences in the individual's wider environment would result in corresponding differences in the individual's thoughts. Take the following example. Suppose a subject, Alex, is related in the right kind of non-representational way to silver—she has a silver bracelet, her mother has a silver tray and a silver tankard, and so on. As such, she acquires the concept *silver*, and comes to believe that silver must be polished to prevent it from tarnishing. Now consider a counterfactual scenario in which there is no silver; a counterfactual scenario in which Alex's bracelet and her mother's tray and tankard are not made of silver, but are instead made of a hypothetical look-alike metal which Alex would be unable to distinguish either practically or theoretically from silver. Let's call the look-alike metal in the counterfactual scenario 'twilver'. In the counterfactual scenario, Alex is by hypothesis related in the right kind of non-representational way not to silver, but to twilver—she has a twilver bracelet, her mother has a twilver tray and a twilver tankard, and so on. This makes it plausible to think that whereas Alex acquires the concept *silver* and comes to believe that silver must be polished to prevent it from tarnishing, counterfactual Alex acquires the concept *twilver*, and comes to believe that twilver must be polished to prevent it from tarnishing. The difference in representational content between the belief Alex has and the belief Alex would have lies in relations to her wider environment—non-representational relations to silver and twilver respectively.

The example of Alex illustrates the fact that an individual's thoughts can depend constitutively on her relations to natural kind properties in her environment. Other Twin Earth thought experiments illustrate the fact that an individual's thoughts can also depend constitutively on her relations to certain socio-linguistic properties, such as how words are standardly used by members of her linguistic community. I will leave the details of these other Twin Earth thought experiments to one side here for the sake of simplicity.[6] This is because Twin Earth thought experiments of all varieties demonstrate a constitutive dependence of one's thoughts on objective properties in one's environment, and this fact

[5] See Putnam (1973) for the original 'Twin Earth' thought experiment as applied to linguistic meaning. See also Kripke (1972). See Burge (1979), (1982) and (1986) for three different applications to mental kinds. Putnam accepts the application to mental kinds in his (1996).

[6] For a recent taxonomy of different kinds of anti-individualism see Sawyer (2011). See also note 4.

alone suffices both to capture the crucial insight of anti-individualism, and to generate the McKinsey paradox, which is the primary focus of the paper.

The Twin Earth thought experiment most often cited in the literature comes from Putnam (1973). In that paper we are introduced to Oscar, an inhabitant of Earth, and his intrinsic physical duplicate, Toscar, who inhabits a planet, Twin Earth, which is exactly the same as Earth except for the fact that the liquid on Twin Earth that occupies the functional role that water does on Earth is not water (H2O), but a look-a-like with a different molecular structure (XYZ). For reasons that parallel the example of Alex, we are invited to conclude that while Oscar possesses the concept *water*, and believes that water is good for quenching thirst, Toscar possesses the concept *twater* and believes that twater is good for quenching thirst. I will draw on this example throughout the paper simply because of its central role in the literature.[7]

3. Privileged access

The thesis of privileged access is the thesis that an individual's epistemic access to the contents of her psychological states is privileged relative to the access others have to the contents of those states. In basic terms—an individual knows what she is thinking in a way in which others do not. To illustrate, suppose Jo believes that the bus is late and is worried that she might miss the start of her lecture. To find out what Jo is thinking, others will have to rely on empirical means. They will have to watch her agitated behaviour, to notice the frequent glances at her watch and her intent focus on the bend in the road around which the bus ought to be coming, to listen to her quiet mumbling about being late for yet another lecture, and so on. For Jo, in contrast, this kind of dependence on empirical means is unnecessary. Jo already knows, before she glances at her watch and hears her worried voice that she believes what she does and fears the perceived consequence. Jo knows her own thoughts not empirically, as others do, but non-empirically—or, as I shall say, broadly a priori.[8]

---

[7] As is now widely recognised, the example is problematic because Oscar and Toscar could not be intrinsic physical duplicates if Oscar's body is made up partly of water, while Toscar's body is made up partly of twater. Since the Twin Earth thought experiments are general in their application, this problem is not systemic, and hence I will treat it as inconsequential for the purposes of the paper.

[8] It is standard in the literature on anti-individualism—as opposed to in the literature on self-knowledge—to classify privileged access as a form of a priori knowledge. It is also standard in the literature on anti-individualism to acknowledge that this classification is not unproblematic. There are clear differences between mathematical and logical knowledge, which provide paradigms of a priori knowledge, and self-knowledge. In

The claim that we have privileged access to the contents of our psychological states is consistent both with our sometimes being wrong about what we think, and with the claim that sometimes others are better placed than we are to determine what we think. The core of the thesis is simply that in many cases, an individual knows what she is thinking in a broadly a priori way.

4. The McKinsey paradox

There is good reason to think first, that our psychological states are anti-individualistically individuated, and second, that we have privileged access to their contents. However, it has been argued that if an individual's psychological states were to depend constitutively on relations to objective properties in her environment, this would preclude the very possibility of her having privileged access to the contents of those states. The apparent inconsistency between anti-individualism and privileged access comes out in two ways, which Martin Davies has called 'the achievement problem' and 'the consequence problem' respectively.[9] According to the achievement problem, if anti-individualism were true, then we could not achieve the kind of privileged access to the contents of our psychological states that we think we have. According to the consequence problem, the hypothesis that we do have privileged access to the contents of our anti-individualistically individuated psychological states apparently has the prima facie absurd consequence that we can have broadly a priori knowledge of the environmental conditions which serve in part to individuate those states. In what follows I will set aside the achievement problem[10] and concentrate exclusively on the consequence problem—also known as the McKinsey paradox, after the introduction of the perceived problem into the literature by Michael McKinsey.[11]

---

addition, some models of self-knowledge specifically liken self-knowledge to a form of empirical knowledge but directed inwards rather than outwards towards the wider environment. I leave these issues to one side here, since they do not affect the force of my argument. By 'broadly a priori', I simply mean 'without recourse to empirical investigation of the individual's environment'.

[9] See Davies (1998).

[10] My reasons for setting aside the achievement problem here are twofold: first, not doing so would take us beyond the scope of the paper; and second, I think the achievement problem has been adequately resolved. See for example Burge (1988), and Heil (1988). See also Sawyer (2002). Of course, an understanding of self-knowledge as contrastive might provide a somewhat different perspective on the resolution, but I don't think it would alter the general spirit of it.

[11] See McKinsey (1991).

The McKinsey paradox brings to the fore an apparent inconsistency between three theses, each of which appears to be true. The three theses are as follows:

(T1) An individual has privileged access to, and hence broadly a priori knowledge of, the contents of her own psychological states.
(T2) The contents of an individual's psychological states depend essentially on relations between her and objective properties in her environment.
(T3) An individual cannot have broadly a priori knowledge of objective properties in her environment.

The apparent inconsistency between these theses can be illustrated by means of an example. Suppose I reflect both on my thoughts and also on the thesis of anti-individualism, then I can reason as follows:

(W1) I am thinking that water quenches thirst.
(W2) If I am thinking that water quenches thirst, then I am related to water (either directly or indirectly).[12]
Therefore:
(W3) I am related to water (either directly or indirectly).

I can know (W1) broadly a priori because this is an unproblematic case in which I have privileged access to the content of my thought. This is an implication of (T1). I can know (W2) broadly a priori because the fact that psychological states depend for their individuation on objective properties in one's environment is a philosophical thesis. This is an implication of the broadly a priori status of T2. But (W3) follows directly from (W1) and (W2). Consequently, it appears that I can have broadly a priori knowledge of (W3), and hence broadly a priori knowledge of an objective property in my environment. This contradicts (T3).

The McKinsey paradox has generated a significant body of literature. For present purposes I set aside responses which reject either (T1), the thesis of privileged access, or (T2),

[12] I would be related to it directly if I had interacted with it myself; I would be related to it indirectly if I had not interacted with it myself but there was a chain of communication connecting me to someone who had interacted with it directly.

the thesis of anti-individualism.[13] Since the primary aim of the paper is to show how we can have privileged access to our anti-individualistically individuated psychological states, responses that reject either (T1) or (T2) are not to the point. But the primary aim of the paper is not merely to show how we can have privileged access to our anti-individualistically individuated psychological states, but to do so without the unwanted implication that we thereby have broadly a priori knowledge of our environment. Given this, I also set aside responses which reject (T3).[14]

My focus, then, will be on responses that aim to preserve the truth of all three theses: (T1), (T2) and (T3). Such responses fall into three broad camps. According to the first, knowledge is not closed under known logical entailment, and hence, applying this in the context of the McKinsey paradox, broadly a priori knowledge of (W1) and (W2) does not imply broadly a priori knowledge of (W3). According to the second, although knowledge is closed under known logical entailment, the warrant for (W1) and (W2) fails to transmit to (W3), which means that (W3) cannot be known broadly a priori *on the basis of* broadly a priori knowledge of (W1) and (W2). According to the third, broadly a priori knowledge of (W2) is unavailable either because (W2) is false, or because, if true, it cannot be known broadly a priori. Let me offer a few brief remarks about each response in turn.

The first response begins with the claim that knowledge is not closed under known logical entailment.[15] Fred Dretske provides a number of examples which, he maintains, demonstrate the failure of the closure principle. According to Dretske, then: (Zebra) I might know on the basis of my visual experience that there's a zebra in front of me, know that if it's a zebra it's not a cleverly-disguised mule, but not know that it's not a cleverly-disguised mule; (Wall) I might know on the basis of my visual experience that the wall in front of me is red, know that if it's red then it's not white and cleverly illuminated by red lighting, but not know that it's not white and cleverly illuminated by red lighting; (Hands) I might know on the basis of my visual experience that I have hands, know that if I have hands I'm not a brain-in-a-vat, but not know that I'm not a brain-in-a vat[16]; and so on. If Dretske were right, and the application to the McKinsey paradox were legitimate, it would follow that broadly a priori

[13] I do not know of anyone who explicitly rejects (T1). Segal (2000) provides a clear rejection of (T2), partly, although by no means entirely, on the grounds that it is inconsistent with (T1).

[14] See Sawyer (1998), (2001) and (2006). See also Warfield (1998).

[15] Dretske rejects the closure principle in his (1970), but see also his (2004). See also Nozick (1981). So far as I know, no-one has endorsed the denial of closure specifically as a response to the McKinsey Paradox.

[16] The reference in (Hands) is to Moore's proof of an external world. See Moore (1939).

knowledge of (W1) and (W2) would not imply broadly a priori knowledge of (W3). This would resolve the McKinsey paradox in such a way as to preserve all of (T1), (T2), and (T3).

There are, however, two concerns with the response. First, it is not clear that the examples really do support the claim that knowledge is not closed under known entailment. Few philosophers have been willing to accept the view, and there are other interpretations of the examples which undermine it.[17] Second, the application to the McKinsey paradox is in any case problematic because it involves a shift in application from perceptual knowledge to broadly a priori knowledge. This is problematic because the intuitive appeal behind the original examples trades on the fact that the visual experience that supports the initial claim in each case is inadequate to support the claim which is known to be entailed by it; but because (W1) is known broadly a priori, there is no evidence to cite which might plausibly support (W1) but not (W3). This leaves us with no general explanation for why closure should fail in such a case.[18]

The second response starts from the claim that restricting knowledge by inference does not require us to reject the closure principle. Accordingly, it upholds the closure principle but maintains nonetheless that the warrant for (W1) and (W2) fails to transmit to (W3).[19] The core idea behind transmission failure is that an argument can be valid, and have both warranted premises and a warranted conclusion, and yet the warrant for the conclusion be such that it is not *derived from* the warrant for the premises—which is to say that the premises provide no *first time* warrant for the conclusion. Transmission, then, is a matter of the structural relations between warrants rather than a matter of their availability. The examples that Dretske takes to demonstrate the failure of the closure principle, Martin Davies and Crispin Wright think are best understood in terms of transmission failure. If this were right, and the application to the McKinsey paradox were legitimate, it would follow that (W1)

---

[17] For reasons not to reject closure see for example Hawthorne (2004).

[18] Two questions arise. First, there is the question of how privileged access to one's own psychological states could yield knowledge if it is not based on evidence. I will not address this question here. Second, there is the question of whether closure might fail not because the evidence that supports a premise is insufficient to support a known consequence of it, but because the known consequence must itself be presupposed in order for the evidence to count as evidence. This way of understanding the issue might provide the basis for an extension from the original, empirical examples to the current, a priori example. The presuppositional status of the conclusion (W3) is addressed in the context of the warrant transmission-failure response which I consider next.

[19] See for example Davies (1998), (2000), (2003) and (2004), and Wright (2000), (2003) and (2011). 'Warrant' is a term of art and different philosophers understand it differently. For my own view, warrant is a positive epistemic property that subsumes both justifications and entitlements. See Burge (1993).

and (W2) would not provide a *first time* warrant for (W3), and hence that (W3) could not be known broadly a priori *on the basis of* broadly a priori knowledge of (W1) and (W2).

But, again, there are problems.[20] First, the claim that the arguments underlying (Zebra), (Wall) and (Hands) suffer from transmission failure depends upon a controversial assumption about the epistemic relation between perceptual experience and perceptual belief. According to the transmission failure response: in (Zebra), my perceptual experience as of a zebra warrants my belief that there's a zebra in front of me only if I am already warranted in believing it's not a cleverly-disguised mule; in (Wall), my perceptual experience as of a red wall warrants my belief that the wall is red only if I am already warranted in believing it's not white and cleverly illuminated by red lighting; in (Hands), my perceptual experience as of hands warrants my belief that I have hands only if I am already warranted in believing I'm not a brain-in-a-vat, and so on. We can agree that perceptual experience warrants perceptual beliefs only if certain background conditions obtain.[21] But the claim that perceptual experience warrants perceptual beliefs only if we are warranted in believing that those background conditions obtain is a much stronger claim, and many philosophers rightly reject it.[22] We can also agree that the warrant that perceptual experience confers on one's perceptual beliefs may be undermined by, for instance, doubts about the honesty of the zoo, doubts about the lighting conditions, doubts about the external world, and so on; but in the absence of such doubts, no prior positive warrant is needed for one's perceptual experience to warrant one's perceptual beliefs. It may be that the beliefs which must be warranted in order for perceptual experience to warrant perceptual beliefs are more general in character than the examples just given suggest—that in (Zebra) one must already be warranted in believing that you're not subject to an elaborate hoax, that in (Wall) one must already be warranted in believing that the lighting conditions are normal, and that in (Hands) one must already be warranted in believing that there is an external world—; but the basic worry remains. If this is right, the examples cannot be understood in terms of transmission failure. The second problem with the transmission failure response is that it is committed to the closure principle and hence committed to the claim that if (W1) and (W2) are known a priori, then so is (W3). Of course, (W3) is known a priori not on the basis of a priori knowledge of (W1) and (W2),

[20] I have argued against the transmission failure response in more depth in my (2006).

[21] Exactly what those background conditions are, of course, is a further question.

[22] See for example Pryor (2000) and Burge (2003).

but antecedently—but it is still hard to see how this constitutes a *solution* to the McKinsey paradox. The claim that (W3) is known a priori constitutes a rejection of (T3).

The third response maintains that broadly a priori knowledge of (W2) is unavailable either because (W2) is false, or because, if true, it cannot be known broadly a priori.[23] There is an element of truth in this response—there is indeed no straightforward connection between possession of a concept and the existence of entities that fall within its extension. According to the anti-individualist, representation in thought depends on non-representational relations to objective properties in one's wider reality. However, it is clearly possible to possess a concept in the absence of entities to which that concept applies.[24] This is consistent with an anti-individualist understanding of the nature of representation. The anti-individualist claim is simply that there could be no representation in thought without *some* relations to objective properties in one's environment. This line of thought is sometimes backed by the example of the chemist who possesses the concept *water* despite living in a world with no water. If her world contained hydrogen and oxygen, and she possessed concepts of each of these elements, it is suggested, she could come to possess the concept *water* in virtue of theorizing about hydrogen and oxygen combining in the appropriate ratio. The example appears originally in Burge (1982), and has been cited prolifically since, with almost universal assent. But the example is problematic. To my mind, such a theorizer might come to possess the concept *H_20,* which represents water, but could not come to possess the concept *water*. The example errs in treating *water* as a reducible concept, which it is not. Water may be identical to H_2O, but *water* is not identical to *H_2O*. While compound concepts, such as *H_20,* can be possessed in the absence of their instances, irreducible concepts cannot. Understood correctly, the example does provide support for the claim that a concept can be possessed by an individual who has had no contact with its instances—either direct or indirect—since the chemist in the example could come to possess the concept *H_2O*, which is a natural kind concept but one that lacks application in her world. Possession of a fundamental, irreducible concept, in contrast, depends essentially on causal relations to its instances. This means that which statements of relations between an individual's concepts and objective properties in her environment are true will depend on which concepts are the fundamental, irreducible ones—and this is plausibly not something that we could know a

[23] See for instance Goldberg (2003), although Goldberg's reasoning differs slightly from the reasoning offered here.

[24] For further discussion of an anti-individualist treatment of empty concepts, see Sawyer (2003) and (2004).

priori. I propose to leave this third response to one side for now, and this for three reasons. First, a contrastive account of self-knowledge provides the means to overcome the McKinsey paradox even if (W2) is knowable broadly a priori. Second, a contrastive account of self-knowledge provides a diagnosis of the McKinsey paradox as a part of a general diagnosis that also explains (Zebra), (Wall) and (Hands). Third, the fact that it is possible to possess a concept in the absence of entities to which that concept applies in fact serves to strengthen the contrastivist solution to the McKinsey paradox.

5. Contrastive self-knowledge

According to the traditional account of knowledge, knowledge is a two-place, categorical relation holding between a subject and a proposition known: Ksp—S knows that p. According to the contrastive account of knowledge, knowledge is a three-place, contrastive relation holding between a subject, a proposition and a contrast class: Kspq—S knows that p rather than that q.[25] It is important to note that the contrast class denoted by 'q' is a set of propositions rather than a single proposition; that the contrast class will typically contain some but not all propositions that contrast with p; and that the set of propositions in the contrast class will therefore typically not be equivalent to ~p.[26]

The contrastive account of knowledge was foreshadowed by Dretske when he wrote: 'To know that *x* is *A* is to know that *x* is *A* within a framework of relevant alternatives, *B*, *C*, and *D*. This set of contrasts, together with the fact that *x* is *A*, serves to define what it is that is known when one knows that *x* is *A*' (Dretske, 1970: 1022).[27] This forms part of Dretske's development of the relevant alternatives theory of knowledge. Since the framework of relevant alternatives which, according to Dretske, partly defines what is known will vary depending on the content of what is known, the theory provides a general principle for rejecting closure, and a general understanding of (Zebra), (Wall) and (Hands). According to the relevant alternatives theory, in each case, the framework of relevant alternatives within

[25] For proponents of the view see note 1.

[26] There will be exceptions. It may be that I know that I exist rather than that I don't. Here, the contrast class relative to which I know the proposition that I exist is identical to the negation of the proposition I know. This is a limiting case: the knowledge is infallible, and hence the proposition known is known in contrast to every other proposition with which it is incompatible.

[27] In fact, much of what Dretske says in his (1970) can be understood as supporting a contrastive account of knowledge. Contrastivism also draws on insights that have been thought to favour a contextualist approach to knowledge. For details see Schaffer (2004).

which the first claim is known is a framework of relevant alternatives within which the claim which is known to be entailed is not known. For example, in (Zebra), the framework of relevant alternatives within which I know that there's a zebra in front of me includes alternatives such as there being an elephant in front of me, there being a giraffe in front of me, there being a pantomime horse in front of me, and so on. Broadly speaking, this is because I can discriminate zebras from elephants, giraffes, pantomime horses, and the like. But, crucially, the framework of relevant alternatives within which I know that there's a zebra in front of me is not one within which I can know that there's not a cleverly-disguised mule in front of me, since, again broadly speaking, I cannot discriminate zebras from cleverly-disguised mules (at least not on the basis of vision). The theory entails the rejection of the closure principle—from my knowledge that it's a zebra, I can come to know the negation of any of the relevant alternatives which provide the framework for my knowledge; but I cannot come to know any claim that lies outside that framework, even one that I know is entailed by what I know. In Dretske's terms, the epistemic operator 'S knows that' does not penetrate to the contrast consequences of what S knows (Dretske, 1970: 1017).

But there are reasons to retain closure, and the contrastive account can do so while at the same time accommodating Dretske's crucial insight that knowledge essentially makes reference to sets of contrasts. The difference lies in the role that the sets of contrasts play in the two different theories of knowledge. The relevant alternatives account sees a contrast set as providing a framework against which a single proposition is known—thus retaining the traditional binary account of knowledge. But the contrastive account sees a contrast set not as providing a framework against which a single proposition is known, but as a very part of what is known, thereby replacing the traditional binary account of knowledge by an account of knowledge as a ternary, contrastive relation. The contrastive account of knowledge, then, maintains that each of the statements in (Zebra), (Wall) and (Hands), while true, is elliptical—what look to be statements of a binary relation between a subject and a proposition known, are really elliptical statements of a ternary relation between a subject, a proposition and a contrast class. Here, then, is the contrastivist understanding of the statements in (Zebra):

> (Z1) I know that it's a zebra rather than that it's an elephant, that it's a giraffe, that it's a pantomime horse....

(Z2) If I know that it's a zebra rather than that it's a cleverly-disguised mule, then I know that it's not a cleverly-disguised mule rather than that it is one.[28]

(Z3) I don't know that it's not a cleverly-disguised mule rather than that it is one.

This allows us to make sense of Dretske's claim that I can know that it's a zebra, know that if it's a zebra then it's not a cleverly-disguised mule, and yet not know that it's not a cleverly-disguised mule. But it does so without having to give up on closure. This is because what Dretske understands as a failure of the epistemic operator 'S knows that' to penetrate to the contrast consequences of what S knows, the contrastivist accommodates explicitly by the fact that the conditional (Z2), while true, does not connect in the right way to the initial knowledge claim (Z1). To connect in such a way as to allow an inference to be drawn, (Z2) would have to have (Z1) as its antecedent, which, now the proper structure of the sentences has been exposed, we can see clearly not to be the case. An inference from 'I know that it's a zebra rather than that it's an elephant' and 'if I know that it's a zebra rather than that it's a cleverly-disguised mule, then I know that it's not a cleverly-disguised mule rather than that it is one' to 'I know that it's not a cleverly-disguised mule rather than that it is one' is a plainly invalid inference. The examples Dretske takes to illustrate the failure of the closure principle, then, the contrastivist understands simply as invalid inferences which, misled by the surface-level grammatical structure of the statements, we mistakenly think are valid.[29]

The question is whether the same contrastivist strategy can provide a solution to the McKinsey paradox. One clear difference between the cases is this: a correct understanding of (Zebra), (Wall) and (Hands) depends on a contrastive understanding of perceptual knowledge, but the McKinsey paradox arises from considerations not of perceptual knowledge but of broadly a priori knowledge—specifically, it arises from the claim that we have privileged access to the contents of our anti-individualistically individuated psychological states. This means that in order to solve the McKinsey paradox in contrastive terms, what we need is a contrastive account of *self*-knowledge. There is independent reason to think self-knowledge is

[28] There is a complication in the formulation here. The original linking statement is a statement of knowledge of a conditional, whereas (Z2) is a statement of a conditional between states of knowledge. Nonetheless, (Z2) states a closure principle as understood in contrastive terms. Similarly, we have (Z2)* If I know that it's a zebra rather than that it's an elephant, then I know that it's not an elephant rather than that it is one.

[29] Similarly, what Dretske thinks of as consequences to which the epistemic operator 'S knows that' *does* penetrate will be understood as valid inferences by the contrastivist. This preserves the crucial point of closure, namely that knowledge can be extended by known entailment.

contrastive, although I will not defend the view explicitly here.[30] Rather, I will outline the view, and then, in the next section, argue that it can indeed be used to solve the McKinsey paradox.

The claim that self-knowledge is contrastive entails that the form of self-knowledge is correctly represented as: S knows that she ψs that p rather than that S* Øs that q. Here, S* marks out a contrast in the subject of the first-order attitude, Ø marks out a contrast in the attitude S takes towards the proposition that provides the content of her first-order psychological state, and q marks out a contrast in the proposition to which S takes that attitude. It is important to note that the contrastivity of self-knowledge is properly located not in the *content* of the self-knowledge (which may or may not be contrastive), but in the subject's *state* of self-knowledge. Thus the content of S's state of self-knowledge when she knows that she ψs that p rather than that S* Øs that q is simply: I ψ that p. But when S knows she ψs that p rather than that S* Øs that q, she will be in a state of knowledge that places her in relation not only to the proposition that she ψs that p (the proposition that provides the content of her self-knowledge) but also to a set of contrasting propositions which differ from it along any or all of the three specified dimensions. Since the McKinsey paradox arises as a result of the constitutive relation between the *content* of a subject's psychological state and relations to her environment, in what follows I will focus exclusively on the contrast class marked out by q—the contrast class that concerns the content of a subject's first-order psychological state when she is in a state of self-knowledge. The contrastive account of self-knowledge proposed is, of course, just a specific instance of the more general contrastive account of knowledge: S knows that p rather than that q.

6. Solving the paradox

We are now in a position to see how a contrastive account of self-knowledge can be used to solve the McKinsey paradox. In order to solve the McKinsey paradox, it will be remembered, we need an explanation of how (T1), (T2) and (T3) are compatible. And the reason this stands in need of explanation is that there are broadly a priori arguments that appear to show that (T1) and (T2) imply the negation of (T3). Let us state such an argument explicitly in terms of knowledge attributions and refer to it as (Water):

(Water1) I know broadly a priori that I am thinking that water quenches thirst.

[30] See Sawyer (2014). In that paper, I also argue for the contrastivity of the propositional attitudes.

(Water2) I know broadly a priori that if I am thinking that water quenches thirst, then I am related to water.[31]

Therefore:

(Water3) I know broadly a priori that I am related to water.

What is the contrastivist to say? Just as with (Zebra), (Wall) and (Hands), the contrastive account of knowledge maintains that each of the statements in (Water) is elliptical—what look to be statements of a binary relation between a subject and a proposition known, are really elliptical statements of a ternary relation between a subject, a proposition and a contrast class. The hope, then, is that by understanding self-knowledge as contrastive, and by providing an explicit statement of the contrasts that are typically left implicit, (Water) will be revealed to be an invalid argument with true premises and a false conclusion. We would then have no reason to think (T1), (T2) and (T3) incompatible.

But there is a question as to the correct specification of the relevant contrasts. In (Zebra), (Wall) and (Hands), the relevant contrasts were marked out by what the subject could and could not distinguish. The natural assumption, then, is that the relevant contrasts in (Water) will also be marked out by what the subject can and cannot distinguish. But it is not the subject's ability to distinguish amongst *objects* that needs to be considered here, but her ability to distinguish amongst *concepts*. How might this be understood? Reflection on the Twin Earth thought experiment involving Oscar and Toscar provides us with the following suggestion: perhaps I am able to distinguish the concept *water* from concepts such as *coffee, alcohol, crackers, silver*, and so on, but am unable to distinguish the concept *water* from the concept *twater*. After all, the Twin Earth thought experiment appears to illustrate the possibility of distinct concepts which are phenomenologically indistinguishable—things would seem the same to me whether I had grown up on Earth and possessed the concept *water* or grown up on Twin Earth and possessed the concept *twater*. Perhaps, then, the following argument, which I will henceforth refer to as (Twater), provides an explicit statement of the contrastivist understanding of (Water)[32]:

[31] For ease of exposition, I have dropped the qualifying expression 'either directly or indirectly'.

[32] For ease of exposition, I will limit the possible contrasts to variations in the 'water'-place, since this is the relevant anti-individualistically individuated concept, but the contrasts would, of course, be much more varied than this simplified account accommodates.

(Twater1) I know broadly a priori that I am thinking that water quenches thirst rather than that I'm thinking that alcohol quenches thirst, that I'm thinking that coffee quenches thirst, that I'm thinking that crackers quench thirst...

(Twater2) If I know broadly a priori that I am thinking that water quenches thirst rather than that I'm thinking that twater quenches thirst, then I know broadly a priori that I am related to water rather than that I am related to twater.

Therefore:

(Twater3) I know broadly a priori that I am related to water rather than that I'm related to twater.

If (Twater) provides an explicit statement of the contrastivist understanding of (Water), then we have revealed (Water) to be an invalid argument. However, this is insufficient by itself to solve the McKinsey paradox. In order to solve the McKinsey paradox (Water) must be revealed to be an invalid argument with true premises and a false conclusion. And here we see that (Twater) fails to deliver the right result: while (Twater) is invalid, its conclusion is nonetheless true. Here's why.

It may be true that things would seem the same to me whether I had grown up on Earth and possessed the concept *water* or grown up on Twin Earth and possessed the concept *twater*—it may be true, that is, that there are possible concepts which are phenomenologically indistinguishable from my own. But it doesn't follow from this that I cannot in fact distinguish the concept *water* from the concept *twater*. To think so would be to conflate two different perspectives: that of the *subject* of the Twin Earth thought experiment and that of the *theorist*. From the perspective of the subject of the Twin Earth thought experiment the concept *water* and the concept *twater* are plausibly indistinguishable. But adopting the perspective of the theorist not only makes it possible for me to distinguish the concepts—it depends on my ability to do so. And, crucially, the McKinsey paradox requires me to adopt the perspective of the theorist, because it requires that I understand the Twin Earth thought experiment—it requires that I know that there are constitutive connections between my psychological states and objective properties in my environment. It follows that the McKinsey paradox depends on my ability to distinguish the concept *water* from the concept *twater*, and hence that it cannot be solved by understanding (Water) as (Twater).

The contrastivist needs to identify a concept which I cannot distinguish from the concept *water* even from the perspective of the theorist. Only such a concept could serve as the relevant contrast in the conclusion of (Water), since only such a contrast could make the

conclusion of (Water) false. But what could that concept be? In order to answer this question, we need to look not to Twin Earth, but to Dry Earth.[33] Dry Earth is a planet which is exactly the same as Earth except for the fact that there is no liquid that occupies the functional role that water does on Earth—rather, the inhabitants of Dry Earth suffer from a pervasive and collective hallucination as of a watery substance. Just as on Earth I would possess the concept *water*, and on Twin Earth I would possess the concept *twater*, let us say that on Dry Earth I would possess the concept *dwater*.[34] All three concepts are indistinguishable from the perspective of the subject. But, as we have seen, adopting the perspective of the theorist in the Twin Earth thought experiment requires that I be able to distinguish the concept *water* from the concept *twater*. This prevents the contrastivist from using the concept *twater* as the relevant contrast in the conclusion of (Water). It might be thought that Twin Earth and Dry Earth are relevantly analogous, and hence that if the concept *twater* is one that the theorist must be able to distinguish from the concept *water*, the concept *dwater* must also be one that the theorist must be able to distinguish from the concept *water*. In fact, however, adopting the perspective of the theorist in the Dry Earth thought experiment does *not* require that I be able to distinguish the concept *water* from the concept *dwater*—indeed, distinguishing the concepts remains impossible. As such, the concept *dwater* is exactly the concept the contrastivist needs to identify as providing the contrast in the conclusion of (Water).

Why is there this difference? Despite their superficial similarity, the Twin Earth thought experiment and the Dry Earth thought experiment are fundamentally different. The expressions 'Twin Earth', 'twater' and 'the concept *twater'* should be treated with care. Take the expression 'the concept *twater'*, for example. It is not an expression that refers to a single concept—by hypothesis it is an expression that aims to refer to a concept that we do not possess. As such, the expression 'the concept *twater'* functions like a variable expression that ranges over all the concepts that are concepts of kinds that are indistinguishable from the kind that we assume our concept *water* refers to. This is the sense in which the expression has an indexical component. And it is this indexical component that ensures that we can distinguish between our own concept and standard twin concepts from the theorist's perspective. The expressions 'Twin Earth', 'twater' and 'the concept *twater*' shift reference depending on the context of utterance. This means that I can know a priori that I am not on Twin Earth, *no*

[33] The example is introduced by Boghossian in his (1997).

[34] What I say here is consistent both with the claim that inhabitants of Dry Earth would fail to express a concept by their use of the term 'water', and with the claim that inhabitants of Dry Earth would possess a concept that fails to refer. For arguments that favour the second option see Sawyer (2003) and (2004).

matter what planet I am actually on. This is analogous to knowing a priori that I am here and not there, no matter where I actually am. Similarly, I can know a priori that I am not related to twater, *no matter what substance I am actually related to*; and I can know a priori that my term 'water' does not express the concept *twater*, *no matter what concept my term expresses.*

The Dry Earth thought experiment, in contrast, presents a situation where things seem the same to me even though my term fails to refer to an objective property in my environment. This can be understood, in contrast to the Twin Earth thought experiment, as a standard sceptical scenario. This means that for all I know, I am on Dry Earth; for all I know, my term 'water' does not refer to an objective property in my environment; and for all I know, my term 'water' expresses the concept *dwater*—the concept *water* and the concept *dwater* may well be the very same concept. I cannot distinguish the concept *water* from the concept *dwater* on the basis of perception, because it is consistent with my perceptual experience that any one of my concepts that I take to refer in fact fails to do so, and hence consistent with my perceptual experience that my term 'water' expresses the concept *dwater*. And I cannot distinguish the concepts a priori, because anti-individualism maintains merely that there could be no representation in thought without *some* relations to objective properties in one's environment, and hence doesn't guarantee that any one of my concepts that I take to refer to an objective property in my environment in fact does so.

I propose, then, that the following argument, which I will henceforth refer to as (Dwater), provides an explicit statement of the contrastivist understanding of (Water):

> (Dwater1) I know broadly a priori that I am thinking that water quenches thirst rather than that I'm thinking that alcohol quenches thirst, that I'm thinking that coffee quenches thirst, that I'm thinking that crackers quench thirst, that I'm thinking that twater quenches thirst...
>
> (Dwater2) I know broadly a priori that if I am thinking that water quenches thirst rather than that I'm thinking that dwater quenches thirst, then I know broadly a priori that I am related to water rather than that my concept fails to refer to an objective property in my environment.
>
> Therefore:
>
> (Dwater3) I know broadly a priori that I am related to water rather than that my concept fails to refer to an objective property in my environment.

(Dwater) is an invalid argument with true premises and a false conclusion, which is exactly what was needed. Since (Dwater) states explicitly the contrastivist understanding of (Water), (Water) is also, according to this contrastivist proposal, an invalid argument with true premises and a false conclusion. Consequently, (Water) cannot be used to show that (T1), (T2) and (T3) are inconsistent. This solves the McKinsey paradox.

An understanding of self-knowledge as contrastive, then, shows how we can have privileged access to our anti-individualistically individuated psychological states without the unwanted implication that we thereby have broadly a priori knowledge of our environment. As such, it provides a solution to the McKinsey paradox. It also clarifies the extent to which we can have privileged access to our anti-individualistically individuated psychological states, and shows why such privileged access cannot be used to refute external world scepticism.[35]

References

Boghossian, P. (1997) 'What the Externalist Can Know A Priori,' *Proceedings of the Aristotelian Society* 97: 161-75.

Burge, T. (1979) 'Individualism and the Mental,' *MidWest Studies in Philosophy* 4: 73-121.

——— (1982) 'Other Bodies,' in A. Woodfield (ed.) *Thought and Object: Essays on Intentionality* (Oxford: Oxford University Press).

——— (1986) 'Intellectual Norms and Foundations of Mind,' *Journal of Philosophy* 83: 697-720.

——— (1988) 'Individualism and Self-Knowledge,' *Journal of Philosophy* 85: 649-63.

——— (1993) 'Content Preservation,' *Philosophical Review* 102: 457-88.

——— (2003) 'Perceptual Entitlement,' *Philosophy and Phenomenological Research* 67: 503-48.

Davies, M. (1998) 'Externalism, Architecturalism, and Epistemic Warrant,' in C. MacDonald, B. Smith & C. Wright (eds) *Knowing Our Own Minds* (Oxford: Oxford University Press).

[35] The contrastivist solution to the McKinsey paradox presented in this paper has more in common with the solution I proposed in my (1999), (2002) and (2006) than might at first appear. Particularly in the last of those papers, I stressed that privileged access to one's anti-individualistically individuated psychological states could yield a certain amount of a priori knowledge of one's environment, but that the knowledge would not suffice to defeat the external world sceptic.

——— (2000) ‘Externalism and Armchair Knowledge,’ in P. Boghossian and C. Peacocke (eds) *New Essays on the A Priori* (Oxford: Oxford University Press).

——— (2003) ‘The Problem of Armchair Knowledge,’ in S. Nuccetelli, (ed.) *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge Mass.: MIT Press).

——— (2004) ‘Epistemic Entitlement, Warrant Transmission and Easy Knowledge,’ *Proceedings of the Aristotelian Society, Supplementary Volume* 78: 213-45.

Dretske, F. (1970) ‘Epistemic Operators,’ *Journal of Philosophy* 67: 1007-23.

——— (2004) ‘The Case Against Closure,’ in M. Steup & E. Sosa (eds) *Contemporary Debates in Epistemology* (Oxford: Blackwell Publishers).

Goldberg, S. (2003) ‘On Our Alleged A Priori Knowledge that Water Exists,’ *Analysis* 63: 38-41.

Hawthorne, J. (2004) ‘The Case For Closure,’ in M. Steup & E. Sosa (eds) *Contemporary Debates in Epistemology* (Oxford: Blackwell Publishers).

Heil, J. (1988) ‘Privileged Access,’ *Mind*, 97: 238-51.

Karjalainen, A. & Morton, A. (2003) ‘Contrastive Knowledge,’ *Philosophical Explorations* VI: 74-89.

Kripke, S. (1972) *Naming and Necessity* (Oxford: Basil Blackwell Ltd.).

Lewis, D. (1996) ‘Elusive Knowledge,’ *Australasian Journal of Philosophy* 74: 549-67.

McKinsey, M. (1991) ‘Anti-Individualism and Privileged Access,’ *Analysis* 51: 9-16.

Moore, G. E. (1939) ‘Proof of an External World,’ *Proceedings of the British Academy* 25: 273-300.

Morton, A. & Karjalainen, A. (2008) ‘Contrastivity and Indistinguishability,’ *Social Epistemology* 22: 271-280.

Nozick, R. (1981) *Philosophical Explanations* (Oxford: Oxford University Press).

Pryor, J. (2000) ‘The Skeptic and the Dogmatist,’ *Nous* 34: 517-49.

Putnam, H. (1973) ‘Meaning and Reference,’ *Journal of Philosophy* 70: 699-711.

——— (1996) ‘Introduction’ to Pessin, A. and Goldberg, S. (eds.) *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam’s ‘The Meaning of “Meaning”’* (New York: M. E. Sharpe, Inc.).

Sawyer, S. (1998) ‘Privileged Access to the World,’ *Australasian Journal of Philosophy* 76: 523-33.

——— (2001) ‘The Epistemic Divide,’ *The Southern Journal of Philosophy* 39: 385-401.

——— (2002) ‘In Defence of Burge's Thesis,’ *Philosophical Studies* 107: 109-128.

——— (2003) ‘Sufficient Absences,’ *Analysis* 63: 202-8.

——— (2004) ‘Absences, Presences and Sufficient Conditions,’ Analysis Vol. 64: 354-57.

——— (2006) ‘Externalism, Apriority and Transmission of Warrant,’ in T. Marvan (ed.) *What Determines Content? - The Internalism / Externalism Dispute* (Cambridge: Cambridge Scholars Press)

——— (2011) ‘Internalism and Externalism in Mind,’ in J. Garvey (ed.) *The Continuum Companion to the Philosophy of Mind* (London: Continuum).

——— (2014) ‘Contrastive Self-Knowledge,’ *Social Epistemology* (forthcoming).

Schaffer, J. (2004) ‘From Contextualism to Contrastivism,’ *Philosophical Studies* 119: 73-103.

——— (2005) ‘Contrastive Knowledge,’ in T. Gendler and J. Hawthorne (eds) *Oxford Studies in Epistemology* 1: 235-271.

——— (2007) ‘Closure, Contrast and Answer,’ *Philosophical Studies* 133: 233-255.

——— (2008) ‘The Contrast-Sensitivity of Knowledge Ascriptions,’ *Social Epistemology* 22: 235-245.

Schaffer, J & Knobe, J. (2012) ‘Contrastive Knowledge Surveyed,’ *Noûs* 46: 675-708.

Segal, G. (2000) *A Slim Book About Narrow Content* (Cambridge, Mass.: MIT Press).

Warfield, T. A. (1992) ‘Privileged Self-Knowledge and Externalism are Compatible,’ *Analysis* 52: 232-7.

Wright, C. (2000) ‘Cogency and Question-Begging: Some Reflections on McKinsey’s Paradox and Putnam’s Proof,’ *Philosophical Issues* 10: 140-63.

——— (2003) ‘Some Reflections on the Acquisition of Warrant by Inference,’ in S. Nuccetelli, (ed.) *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge Mass.: MIT Press).

——— (2011) ‘McKinsey One More Time,’ in A. Hatzimoysis (ed.) *Self-Knowledge* (Oxford: Oxford University Press).