The Puzzle of Transparency

Ram Neta
UNC-Chapel Hill

I. Knowing Whether You Believe Something: A Puzzle

How do you know what you believe? Gareth Evans addresses this question in the following famous passage:

"in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the question whether p." (Evans 1982, 225)

From now on, let's refer to the observation that Evans makes here as his "transparency" observation, and let's refer to the phenomenon so observed as the "transparency" of belief. This phenomenon is as puzzling as it is obvious. While it is no doubt true that we *often* respond – and often *ought* to respond – to questions of the form "do you think that p?" by considering whether it is true that p, this fact about how we do and ought to respond to such questions is puzzling, and such puzzlement can be articulated as follows:

"What right have I to think that my reflection on the reason in favor of p (which is one subject-matter) has anything to do with the question of what my actual belief about p is (which is quite a different subject-matter)? Without a reply to this challenge, I don't have any right to answer the question that asks what my belief [about, e.g., whether it will rain] is by reflection on the reasons in favor of an answer concerning the state of the weather." (Moran 2003, 405)

So here is what I will henceforth call "the puzzle of transparency". On the one hand, we have the phenomenon of transparency, which is an obvious fact: the fact that we normally do and should attend to our evidence whether p in order to answer the question whether we believe that p. On the other hand, it's unclear how this seemingly obvious fact could possibly be true: how could evidence as to whether p tell us anything about our states of mind, if p itself is not about our states of mind? This puzzle calls

for some philosophy: we need to find a way of understanding the seemingly obvious fact so as to avoid the puzzle of transparency. How to do this?

Before I try to answer this question, let me briefly allay a worry that may arise about the puzzle that I've articulated. Some philosophers are inclined to think that, in the situations that exhibit the transparency that Evans and Moran describe, when I ask you "Do you think there's going to be a third world war?", what I'm *really* asking is rather something like this: Given the evidence currently at your disposal, are you about to start believing that there is going to be a third world war? But this proposal cannot be correct: when I ask you if you think that there's going to be a third world war, I am not asking you to make a prediction about what you will believe: a prediction can only be made when the predicted event hasn't yet occurred, but in this case, as soon as you answer my question, you also believe. So there's no time lag between your putative "prediction" and that which you are, on this proposal, predicting. Perhaps what I'm really asking is rather something like this then: Given the evidence currently at your disposal, are you disposed to believe that there is going to be a third world war? If this is what I'm asking, then it raises a slightly different puzzle: how can I gain knowledge of what dispositions I have by considering geopolitical evidence concerning whether there will be a third world war? Can I do so by knowing how I respond (even if only in thought) to such evidence? Perhaps I can, but then how do I know my own mental responses to that evidence, after being asked the question? There must be some explanation of how I know my mental responses to my own evidence, an explanation that needn't involve my observing my behavior (since perhaps I don't engage in any observable behavior), but that still involves my gaining such knowledge by considering (say) the geopolitical evidence for or against the occurrence of a third world war. And so the puzzle of transparency can still be formulated even if we choose to reinterpret the question "Do you believe there will be a third world war?" in the way just suggested.

Moran offers one suggestion, immediately after the passage quote above:

"I *would* have a right to answer that my reflection on the reasons in favor of rain provided me with an answer to the question of what my belief about the rain is, if I could assume that what my belief here is was something determined, by the conclusion of my reflection on those reasons."

Thus, if I am entitled to assume that my belief about whether it will rain is formed by my reflection on the reasons that bear on whether it will rain, I will then also be entitled to assume that whatever conclusion I arrive at in reflecting on those reasons is a conclusion that I will believe to be true. It is thus, Moran concludes, my entitlement to the latter assumption – which

derives from my entitlement to the former assumption – that allows me to know what beliefs I hold by reflecting on the evidence for the truth of their contents.

Moran's view invites two questions. The first question has to do with the connection between the first thing that he claims I am entitled to assume, viz., that my belief about whether it will rain is formed by my reflection on the reasons that bear on whether it will rain, and the second thing that he claims I am entitled to assume, viz., that whatever conclusion I arrive at in reflecting on those reasons is a conclusion that I will believe to be true. And the second question has to do solely with the grounds for his claim that I am entitled to assume the first of those two things. Let me take a moment to articulate each of these two questions.

Moran says that what entitles me to assume that the conclusion I arrive at by reflecting on my reasons is a conclusion I believe to be true. To see just what this connection is, we would have to make clear what it is to *arrive* at a conclusion as to whether p, and to do so by *reflecting* on reasons that bear on whether p. If arriving at a conclusion as to whether p by reflecting on reasons that bear on whether p is simply a matter of drawing the conclusion that p from some considerations that bear on whether p, then it's not clear that there is much substance in this assumption at all: the assumption simply amounts to the claim that, when I draw the conclusion that p from some premises that are relevant to whether p, I thereby believe that p. But it's not clear what it could be to draw a conclusion from some premises if it doesn't involve at least believing the conclusion on the basis of the premises. Thus, what entitles me to assume that *the conclusion I arrive at by reflecting on my reasons is a conclusion I believe to be true* is nothing more than my understanding the content of the assumption itself: the assumption is not something that could rationally be denied by anyone who understood it, no matter what else they might take to be true.

But now it appears that Moran has solution to our puzzle does not involve any link between the two things that he claims we are entitled to assume. Our puzzle was this: how can consideration of the reasons for or against p tell us what we believe? Moran's solution was to say that we are entitled to assume that whatever conclusion we draw from consideration of the reasons for or against p is a conclusion that we believe to be true. And it is plausible that we are entitled to assume this, because what we thereby assume is nothing more than what is implicit in our concept of *drawing a conclusion from some considerations*. But our understanding of that concept seems to have nothing to do with our assumption that we are entitled to assume that our beliefs are formed by our reflection on our reasons for them.

There is a second question, however, about Moran's solution to the puzzle, and it is about our entitlement to assume that our beliefs are formed by reflecting on the reasons for or against their truth. Moran claims that we are entitled to assume this, and his claim is not implausible. But it cannot be a primitive, inexplicable fact about us that we are entitled to assume this. There must be some feature of us that furnishes us with this entitlement, and in virtue of which the entitlement may be stronger for some agents than for others, and stronger for some topics than for others, or stronger in some situations than in others. In short, though Moran may very well be right to claim that we enjoy such an entitlement, without explaining what provides us with this entitlement, or why it is as strong as it is, when it is, his explanation of the transparency puzzle is superficial and ad hoc.

In this paper, I will articulate and defend what I take to be a successful version of Moran's solution to the puzzle of transparency – a version that answers both of the two questions that I've just raised. In order to do so, however, it will be useful to begin by critically examining an alternative to Moran's solution, provided by Alex Byrne.[1]

II. Byrne's Proposed Solution to the Puzzle

According to Byrne, when our considerations of reasons for or against p leads us to believe that p, we then typically acquire knowledge that we believe that p by making an inference of the form:

P

I believe that P.

Of course, Byrne admits, such an inference is not valid, nor does the truth of the premise in general increase the likelihood of the conclusion being true. Nonetheless, when we make such an inference, we typically acquire knowledge that the conclusion is true, and that is because we know a priori that inferring the conclusion by means of this inference is a perfectly reliable way of coming to believe the conclusion: we cannot infer the conclusion from the premise unless we believe the premise – which is just to say, unless the conclusion is true – and that is all knowable a priori, simply by reflecting on what it is to *infer a conclusion from a premise*. Since we can know a priori that any conclusion reached by making an inference of the form above is true, we can achieve knowledge of any such conclusion by making that inference. Byrne thus seems to offer a solution to our original puzzle.

[1] See Byrne 2011.

Of course, some epistemologists will be tempted to reject Byrne's proposed solution on the grounds that we allegedly cannot gain knowledge of the truth of some proposition by inferring that proposition from a false premise. But this is not my objection to Byrne's proposal: I'm happy to grant Byrne that, if I form a belief in a way that I know a priori is completely reliable, then I can know the belief so formed to be true. So what, then, could be the problem with Byrne's proposed solution?

Recall that, for Byrne's proposed solution to work, it must explain how it is that, *quite generally*, my consideration of reasons for or against p enables me to know whether I believe that p. So the sorts of inferences that Byrne describes must be inferences that I make routinely. Indeed, on Byrne's own view, I must make them just as routinely as my self-knowledge exhibits the transparency phenomenon, since the puzzle about transparency arises for every instance of that phenomenon. Of course, not all of my self-knowledge exhibits the transparency phenomenon in this way: for instance, when I meditate, I do not figure out what thoughts or feelings or sensations are passing through my consciousness by considering any extra-mental matters of fact. But even if not all of my self-knowledge exhibits the phenomenon of transparency, much of it does – including my knowledge of whether I believe a proposition, whether I intend to do a particular thing, whether I perceive a particular object, and so on. Thus, on Byrne's view, it is by means of perfectly analogous inferences that I routinely come to know my perceptions, intentions, and other attitudes as well. And so, if Byrne's proposed solution works, then I *typically* and *rationally* make inferences from premises about the extra-mental world to conclusions about my beliefs or other mental states.

As ingenious as Byrne's proposed solution is, it will not work. There are at least two serious problems with it, both of which have to do with the nature of inference. In the next section, I will articulate these problems.

III. Two Features of Inference, and why Byrne's Proposed Solution to Our Puzzle Cannot Do What It Promises To

My first objection to Byrne's proposed solution doesn't show that the solution is wrong, but does show that it doesn't offer the benefits that Byrne claims for it. The objection has to do with the fact that much of its appeal as a solution to our puzzle derives from its generality. According to Byrne: inferences of the form that he describes not only account for the fact that we can answer questions about whether we believe that p by considering the evidence concerning p, but it can also account for the fact that we can answer questions about whether we enjoy an experience as of such-and-such objects by considering whether such-and-such objects are

before us, and it can also account for the fact that we can answer questions about whether we intend to F by considering the reasons for F'ing, and so on. In short, Byrne recommends his proposed solution to our puzzle on the grounds that it explains how we achieve self-knowledge by means of outwardly directed attention in all the various cases that exhibit the transparency phenomenon. But now I will argue that it cannot explain the whole range of cases of transparency. To see this, let's consider how Evans's observation might apply to the case of *knowing why* you believe something, rather than *knowing whether* you believe it. Imagine that Evans had written the following

In making a self-ascription of one's reasons, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me '*Why* do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'What indicates that there will be a third world war?' I get myself in a position to answer the question why I believe that p by putting into operation whatever procedure I have for answering the question what indicates that p.

The preceding paragraph, obviously adapted from the passage quoted at the beginning of this paper, makes a point that is just as plausible: just as it is typically rational for us to answer the question *whether we believe that p* by considering whether p, it is also typically rational for us to answer the question *why we believe that p* by considering what indicates p. But why would it be rational for us to answer the question why we believe p by considering what indicates p? This puzzle is analogous to our original puzzle: both are puzzles of transparency. But how can Byrne's proposed solution to our original puzzle generalize to explain this new puzzle of transparency? Could Byrne propose that we gain knowledge of why we believe that p by making an inference of the form:

E indicates that p.

I believe that p on the grounds that E.

No: the hypothesis that we make such inferences cannot explain why we can typically rationally answer the question "Why do you believe that p?" by considering what indicates that p. That's because we can accept that a particular piece of evidence E indicates that p, while consistently granting that E's support for p is defeated – and if we do the latter, then we may grant the premise without drawing the conclusion. And so inferences of the form just sketched do not enjoy the apriori demonstrable reliability of the inferences by appeal to which Byrne proposes to solve our puzzle about how we can know what we believe by considering what is true.

Could Byrne then propose that we gain knowledge of why we believe that p by making an inference of the form:

E indicates that p, and is not defeated.

I believe that p on the grounds that E.

Once again, no: the hypothesis that we make such inferences cannot explain why we can typically rationally answer the question "Why do you believe that p?" by considering what indicates that p. That's because we can accept that a particular piece of evidence E indicates that p, and is not defeated, while consistently granting that E is not strong enough to demand belief in p, and also consistently refusing to take a position as to whether p is true. And again, if we do the latter, then we may grant the premise without drawing the conclusion. And so, once again, inferences of the form just sketched do not enjoy the a priori demonstrable reliability of the inferences by appeal to which Byrne proposes to solve our puzzle about how we can know what we believe by considering what is true.

Could Byrne then propose that we gain knowledge of why we believe that p by making an inference of the form:

E conclusively indicates that p, and is not defeated.

I believe that p on the grounds that E.

Yet again, the answer is no: the hypothesis that we make such inferences cannot explain why we can typically rationally answer the question "Why do you believe that p?" by considering what indicates that p. And that's because we can rationally answer the question "Why do you believe that p?" by considering what indicates that p, *even when* we do not regard ourselves as having conclusive indication concerning the truth of p. So even if we sometimes do make the inference above, and even if our making it can explain why it is sometimes rational for us to answer the question why we believe that p by considering what indicates that p, our making such an inference cannot explain why it is rational for us to answer the question in this way even when we don't believe there to be conclusive indication that p. And so, yet again, this proposal fails.

Could Byrne then propose that we gain knowledge of why we believe that p by making an inference of the form:

My total evidence T supports p strongly enough to make it credible that p.

I believe that p on the grounds contained in T.

The answer to this question seems to depend on what "credible" means. If the credibility of a proposition on one's total evidence amounts to one's

total evidence making belief in the proposition mandatory, then this proposal is too strong: we can know why we believe that p even if we don't it to be mandatory for us to believe that p. If the credibility of a proposition on one's total evidence amounts to one's total evidence making belief in the proposition permissible, then the proposal is too weak: we can regard belief in a proposition as permissible even if we don't believe the proposition ourselves. Finally, if the credibility of a proposition on one's total evidence amounts to one's total evidence making belief in the proposition an all-things-considered *good idea*, then this proposal is once again too weak, since we can regard belief in a proposition as an all-things-considered good idea even if we recognize ourselves as unable to believe the proposition ourselves.

We have tried a number of ways to extend Byrne's proposal to explain how it's possible for us to know whether we believe that p by considering whether p is true, so that it could also explain how it's possible for us to know why we believe that p by considering what indicates that p is true. But no such extension of Byrne's proposal has any hope of success. I conclude that much of the appeal of Byrne's proposal, viz., its generality, is illusory.

Of course this first objection to Byrne's proposed solution is still consistent with its truth, even if not with its generality. But my second objection to Byrne's proposal shows that it cannot even be true. This second objection is *not* the often voiced objection that inferences of the form Byrne describes cannot be rational[2]: indeed, I think that, *in the right circumstances*, they can be rational. Rather, my objection is that the transitions that he describes cannot *typically* be inferences at all. Not all transitions from one set of beliefs to another constitute an inference – not even if the transition necessarily results in beliefs that are known a priori to be true. To see this, consider a logician who is caused to prove a new (but boring) theorem every time she accepts a premise of the form "I see something green". This is a strange causal process, to be sure, but the logician is in a position to know *a priori* that the beliefs formed at the end of this causal process will all be true: they are provable theorems, after all! So the logician knows the theorems that she proves, and her belief in those theorems is caused by a process that she knows a priori will result in beliefs that are true. But is the causal process by means of which she forms these beliefs an *inference*? Barring some extraordinary elaboration of the case, the answer is no: not all causal processes that take beliefs as input and deliver beliefs are output are inferences. For an agent to make an inference, rather than simply undergoing a causal process in which her beliefs are altered, the agent must satisfy something like what Boghossian 2012 calls "the taking condition": the causal process must be one that the agent undergoes *in virtue of thinking*

[2] See Boyle 2011 and Barnett 2016.

that the premise justifies the conclusion.[3] But an agent who had this thought about the inference from P to "I believe that P" would either be suffering from a severe confusion, or take it that P could not be true without her believing it, or both: an agent cannot rationally take the inference from P to "I believe that P" to be a good inference unless she rationally believes that P could not be true without her believing it: and with respect to how many propositions could someone rationally believe that? An agent who believes it about all the propositions in her belief-set is decidedly *not* the ordinary, rational self-ascriber that Evans means to be portraying. Indeed, if I *ever* manage to form a belief of the form "I believe that P" by inference from the corresponding premise that P, it will only be because either I suffer from some confusion that prevents me from accurately tracking my own reasoning, or else I take myself to be omniscient with respect to the relevant facts. But neither of these two scenarios obtains very often. So, on the rare occasion when I can make the inference from P to "I believe that P", that inference will typically be irrational – either the result of my losing track of the inference I'm making, or else the result of my taking myself to be omniscient with respect to P-relevant facts, and the latter is typically irrational. That is my second objection to Byrne's proposed solution to our puzzle: his solution can explain why Evans's procedure is *typically rational* only for those extraordinary agents who rationally believe that a proposition cannot be true unless they believe it to be true, and it is can explain why Evans's procedure is even *sometimes rational for you* only for those extraordinary propositions which are such that they cannot be true without your believing them. And this shows that Byrne's proposed solution is not just inadequately defended (as my first objection showed). This shows that Byrne's proposed solution is also false.

Our puzzle, recall, was this: On the one hand, we normally do and should attend to our evidence whether p in order to answer the question whether we believe that p. On the other hand, it's unclear how this seemingly obvious fact could possibly be true: how could evidence as to whether p tell us anything about our states of mind, if p itself is not about our states of mind? Byrne doesn't successfully address this puzzle, and Moran seems not even to try to address it. How, then, can we address it? It is as tempting as it is common to think that the answer to this question must be along the

---

[3] Boghossian states the taking condition as follows: "A transition from some beliefs to a conclusion counts as inference only if the thinker *takes* his conclusion to be *supported* by the presumed truth of those other beliefs." (Boghossian 2012, 4) There are many different ways of specifying this condition, and not all of them are specification on which the condition is necessary for inference. But in Neta forthcoming I argue that one such specification is necessary for inference. Thomson 1965 offers what I believe to be the earliest defense of the taking condition on inference.

following lines: there is a mechanism that detects when you form the belief that p, and then, upon detecting this belief, somehow brings you into possession of the knowledge that you believe that p. All normal humans have such a mechanism, and it detects only our own states of mind – not those of others. The mechanism typically operates with much greater reliability than any or our ordinary perceptual mechanisms, which is why normal humans are so much more reliable when it comes to their own states of mind than when it comes to features of the extra-mental world. Of course, the mechanism has its limits: it tends to produce unrealistically flattering self-attributions of motives, and it fails to detect many of our mental states altogether. But such limits are consistent with its being reliable enough to make our knowledge of our own mental states "privileged" in relation to our knowledge about most other topics. Following Finkelstein 2003, I'll call this view "detectivism".[4]

It is sometimes thought that detectivism is refuted by Shoemaker's argument against the possibility of self-blindness[5], or by Burge's argument that knowledge of your own states of mind is a necessary condition of being subject to rational requirements.[6] But this is not the case. Even if Shoemaker is correct that a rational agent who has the concept of belief must know what she believes, it doesn't follow from this that she cannot know it by means of the detectivist's posited mechanism: Shoemaker's argument leaves it open that knowing it by means of such a detecting mechanism is itself a necessary condition of having the concept of belief, or of being rational, or of satisfying the conjunction of these two conditions. Again, even if Burge is correct that an agent who is subject to rational requirements must know what she believes in order to have the ability to comply with those requirements, it doesn't follow from this that she cannot know it by means of the detectivist's posited mechanism: Burge's argument leaves it completely open exactly how we know what we believe. So neither of these modern transcendental arguments tells against detectivism.[7]

Detectivism can be fleshed out in many ways, depending upon how the detecting mechanism is characterized: what it takes as input, and how it converts that input to output. But however precisely the view is fleshed out, and independently of whether the view is true, it is not clear how the view can fully address our puzzle. Recall that our puzzle was this: On the one hand, we normally do and should attend to our evidence whether p in order to answer the question whether we believe that p. On the other hand, it's

---

[4] Perhaps the most prominent proponent of detectivism (though not under that name) is Armstrong 1968.
[5] Shoemaker 1994
[6] Burge 1996.
[7] I make this point in Neta 2011.

unclear how this seemingly obvious fact could possibly obtain: how could evidence as to whether p tell us anything about our states of mind, if p itself is not about our states of mind? In response to this puzzle, detectivism says: we have a mechanism that works in just the way it would need to work in order to make the seemingly obvious fact obtain: it detects our beliefs, and delivers us knowledge that we have those beliefs. If this counts as a response to our puzzle, then it's not clear what could have been so puzzling in the first place. Compare this "solution" to our puzzle with the following solutions to classic philosophical puzzles: How is free will possible? There is a mechanism that operates to insure that some of our actions are done freely. How is knowledge of the external world possible? There is a mechanism that operates to insure that we enjoy knowledge of the external world. Positing a mechanism that does some task cannot solve a puzzle about how that task is so much as possible: the only solution to such a puzzle has to address the source of the appearance of impossibility.

Recall that our original puzzle was not a puzzle about how it's possible to know what we believe, but rather a puzzle about how we can answer the question what we believe by considering evidence that bears on some issue other than the issue of what we believe. We can simply restate this puzzle within the detectivist framework as follows: how can we know what we believe about an issue by doing what, according to the detectivist, simply amounts to forming a belief about that issue? Of course, forming a belief about some issue will make it the case that *there will be something for us to find out* concerning what we believe about that issue. But how can forming a belief about some issue make it the case that, at the time of being asked what we believe, and thus, prior to considering the evidence and forming a belief about the issue, *there was something for us to have found out* concerning what we believe about that issue?

Suppose that, when someone asks you if you have a headache, their asking that question causes you to have a headache. In that case, when they ask you the question, the correct answer is "yes". But if they are asking you the question in order to learn whether, prior to, or independently of, being asked, you had a headache, then they will not learn what they want to learn by accepting to your truthful answer to their question. Analogously, if someone asks you whether you believe that there will be a third world war, and their asking you this question causes you to consider the evidence and arrive at the conclusion that there will be, then the correct answer is "yes". But if they are asking you the question in order to learn whether prior to, or independently of, being asked, you believed that there would be a third world war, then they will not learn what they want to learn by accepting your truthful answer to their question. Thus, if we accept Evans's observation concerning how we typically and rationally answer questions of the form "Do you believe that p?", it seems that we must also accept the

following conclusion: when such questions are normally asked, the interrogator is not interested whether we hold the belief in question at the moment that the question is being asked, but rather whether we will hold the belief very soon thereafter.

Of course, this conclusion is implausible: if the interrogator were interested in whether we are going to believe that p, why wouldn't she simply ask something of that form? Or ask whether we are willing to accept that p given our current evidence? The distinction between what we are willing to accept and what we already believe is an easy and familiar distinction, and it's explicitly marked in ordinary language often enough that we should be dubious about any view that accuses ordinary speakers of typically eliding the distinction.

But this leaves us with the following puzzle: We normally, rationally, answer questions of the form "Do you believe that p?" – questions concerning what we now believe, and not what we are soon going to believe – by considering the evidence whether p. And yet, when p concerns some non-psychological matter of fact, the evidence whether p seems to indicate nothing about whether we believe that p. So how could the procedure that Evans treats as our normally rational procedure be normally rational? How could it be typically rational to answer a question whether some state of affairs obtains by considering evidence that indicates nothing about the obtaining of that state of affairs?

In order to answer this question, I need to take a detour through the metaphysics of institutional action. After the detour, I will finally be in a position to propose a novel solution to our puzzle.

IV. **Essentially Represented Kinds and Particulars**

There are many different series of physical events that could constitute checkmating an opponent in chess: these events could involve moving a carved piece of wood with your hand, pushing a carved piece of metal with a stick, or pressing buttons on a computer, among countless others. Consider any particular one of these series – say, the series of events that includes that various motions of my arm and fingers around the tallest carved piece of wood sitting on a chessboard on the second floor of my house in North Carolina on October 3, 2017, at 2:43 PM, Eastern Standard Time. What makes it the case that some series of events I've just mentioned constitutes *checkmating my opponent*? In part, it's the fact that this series of events is done *intentionally* – it's not, for instance, a spasm – and it is done *with the intention* of moving a particular chess piece to a particular position on the board, and in part, it's the fact that, once that chess piece occupies

that position on the board, the opponent is in checkmate. But what makes that particular carved piece of wood a *chess piece*, and what makes a particular painted square on a wooden surface a *position on the board*, and what makes a particular configuration of pieces *checkmate* for one of the players? Once again, the answer will have to do with the intentions with which some group of people designed and created these various things or statuses, or the intentions with which another group of people use these things or respond to these statuses, and so on. In short, what makes a particular series of physical events into the event of checkmating an opponent in chess, and what makes particular things into chess pieces or chess positions or checkmate, is the intentions of human beings with respect to these things. Since intentions are representations, it follows that what makes a particular series of things into chess events or chess objects or chess statuses of certain kinds is the way in which those events or objects or statuses are, at some time or other (not necessarily at the time of their creation, nor necessarily at the time of their use) represented. It is in the nature of a chess piece to be *represented as a chess piece of a particular kind*, in the nature of a chess event to be represented *as a chess event of a particular kind*, and in the nature of a chess position to be represented as *a chess position of a particular kind*. We can put this point by saying that chess kinds are *essentially represented* kinds: kinds to which a particular can belong only if the kind is somehow represented.

Of course it doesn't follow from this that whatever is represented as a chess piece must actually be a chess piece. Being represented as a chess piece is merely a *necessary*, but not a *sufficient*, condition of being a chess piece. It also doesn't follow from the claim that chess kinds are essentially represented kinds that *each* chess piece needs to be represented *de re*: all that follows is that something is a chess piece only in so far as it satisfies the conditions for falling into the extension of a general representation – but the representation must, at some point or other, be actual. Molecules could exist even if there had never been representations of molecules, or of anything else, but chess pieces could not exist unless there had been representations of chess pieces *as such*. Of course, none of this is to deny that each chess entity needs to be represented de re in order to belong to its chess kind: in the case of at least some chess events like moving a particular piece to a particular position, it *is* necessary for each such event to be appropriately represented *de re* in order to belong to the particular chess kind to which it belongs – we may say that such particulars do not merely belong to essentially represented kinds, but they are furthermore *essentially represented particulars*. Finally, it doesn't follow from the claim that chess kinds are essentially represented kinds that being represented as a chess piece of a particular kind is necessary for being a chess piece *of that particular kind*. Being represented as a chess piece of some kind is a necessary condition of being a chess piece of some kind, but the representation needn't be wholly accurate in order to constitute the piece as

a piece of its kind. If a particular manufacturer of chess sets becomes confuses and thinks that the "Queen" pieces are not Queens but are rather "Jacks", this does not suffice for them to be Jacks: if they are chess pieces, then they can't be Jacks, since there is no such chess piece. So the representation needed to constitute the pieces as pieces of the kinds that they are need not be a wholly accurate representation: it may be inaccurate, confused, or in various other ways defective, without thereby ceasing to do the metaphysical work of, say, making the carved piece of wood count as a Queen.

I've so far described this metaphysical work in a way that may suggest the following picture: physical reality contains carved pieces of word, painted squares on wooden boards, and series of events involving my arm grasping a particular piece on a particular painted square, and then moving that piece to a different square. Representations of certain kinds then make the carved piece of wood count as a Queen, the wooden board count as a chess board, the painted square count as a particular chess position, and the series of movements of my arm and fingers count as the Queen's moving from one position to another. But things need not be quite this simple. For instance, *precisely which* series of movements of my arm and fingers count as the Queen's moving from one position to another? We may be tempted to reply that it is whichever series of movements is done with the intention of moving the Queen from one position to another – but *precisely which* series of movements is done with that intention? Does it include the motions I make with all of my fingers or just the fingers in contact with the Queen? Does it include the motions made as I reach towards the Queen, or just the motions made while my fingers are in contact with the Queen? Does it include the motions that I make with my elbow, in an effort to balance myself while moving the Queen? My intentions are typically not fine-grained enough to settle these questions: when I intend to move my Queen from one position to another, I typically don't also intend to do it by means of the precise physical motions in which I engage in the course of doing it. This is not to say that some of my motions are *unintentional*, but only that which specific physical motions are part of my intentionally moving the Queen from one position to another is left less than fully determinate by my intentions. Is there anything that makes it fully determinate which specific physical motions are part of my intentionally moving the Queen from one position to another? Maybe there is, but maybe not. If not, then there is no *precise* series of physical motions that counts as my moving the Queen, and so as my checkmating my opponent. In that case, my checkmating my opponent is not identical with any precise sequence of physical motions: the spatio-temporal boundaries of the event may be no more determinate than the spatio-temporal boundaries of, e.g., *lunchtime*. It doesn't follow from this that my checkmating my opponent consists in some *non*-physical motions in addition to a series of physical motions. Checkmating my opponent is an event that may or may not be

identical to a precise series of physical motions, but whether identical or not, it is constituted by that (precisely or vaguely spatio-temporally bounded) series of physical motions, so long as it (the event of checkmating) is represented *as a chess event* of some such kind. We may say that the (precisely or vaguely) spatio-temporally bounded series of physical events is the *matter* from which the event of checkmating is constituted – that series is what the event of checkmating is *made of* – but the representation of that event as a chess event of some kind unifies that matter into a single event with a particular *form*, related to the content of the representation.[8]

I've so far tried to clarify and defend the hypothesis that chess kinds are essentially represented kinds. Now let me add one further point about chess kinds in particular, and essentially represented kinds in general. Suppose that I have two possible moves open to me: I decide to move my queen to g7 or f7, but I haven't yet decided which. At time t1, I pick up my queen and move it in the air towards that corner of the board as I contemplate my options. Then, at time t2, I decide on g7. At time 3, I put my queen down on g7. So moving my queen to g7 was a move that started at t1 and ended at t3. But notice that none of the non-chess events that took place between t1 and t2 constituted its being the case that I was moving my queen to g7. This is true even if causal determinism is true: even if the events that took place between t1 and t2 *causally determined* that I was going to move my queen to g7, they still didn't *constitute* my moving my queen to g7. That's because the event of my moving my queen to g7 is an event that lasted from t1 until t3, and so lasted longer than the series of non-chess events from t1 until t2. So the event of my moving my queen to g7 is an event that begins at t1, even though at least one of the representations by virtue of which that event is one of moving my queen to g7 doesn't occur until t2. The nature of the event is not determinate at the time at which the event begins – it only becomes determinate after the event begins.

I can now summarize the points I've made in this section as follows. Chess kinds are essentially represented, and so for an object or event or status to belong to a particular chess kind requires that chess kinds are actually represented. Furthermore, some particulars of those kinds are essentially represented, and for the particular to exist requires it be represented de re. But at least in some cases, the representations that are needed to make it the case that a particular of a certain kind exists – those representations do not themselves exist until after that particular begins to exist. What do these points have to do with our original puzzle about how we can know whether we believe that p by considering whether p is true?

V. The Solution to Our Puzzle: the Mental Particulars that Generate the Puzzle are Essentially Represented

[8] The argument of this paragraph is adapted from Haugeland 1982.

Evans brought to our attention a particular puzzle concerning our knowledge of whether we believing some proposition: we can gain such knowledge by considering whether the proposition in question is true. In section III, I pointed out that there is an analogous puzzle concerning our knowledge of why we believe some proposition: we can gain such knowledge by considering what indicates the truth of that proposition. In both cases, we gain knowledge of some of our own mental states (beliefs) or acts (inferences) by considering facts that are logically and metaphysically independent of whether we occupy those states or perform those acts. How is this possible?

Before answering this question, let me also remind the reader of one other promissory note that I issued above. In discussing Moran's attempt to solve our puzzle, I granted the claim that in reviewing some considerations concerning some non-psychological topic, we might not only come to draw a conclusion from those considerations, but also come to know that we are drawing a conclusion from those considerations. I asked how we could come to know this latter, and then mentioned that Moran didn't address this question – and didn't address it for reasons that we would eventually discuss. It is now time to discuss those reasons.

Moran is happy to grant that even if detectivism cannot solve the puzzle with which this paper began, some version of detectivism is nonetheless true concerning much of our knowledge of our mental states and processes: I can introspect carefully, and accurately report what sounds I hear, what colors I see, what mental images I enjoy, what I recall of some occasion, and what feelings are conjoined with my recollection. None of these cases exhibit the puzzling behavior that Evans described though: if asked whether I am now having a mental image as of a red square, or whether I now recall my 7th birthday party, or whether I now feel more like listening to Miles Davis or Edith Piaf, I do not typically answer by considering extra-mental facts. The puzzling phenomena with which we began are phenomena that surround only some, but not all, of our knowledge of our own minds. We want to solve the puzzle of how we can acquire such knowledge, when we do acquire it, by considering extra-mental facts.

I now propose to solve our puzzle by appeal to the following hypothesis: the cases that exhibit the transparency phenomenon are cases of mental states or mental acts that have both of the following two features:

(a) They are essentially represented: the agent occupies the mental state or performs the mental act *only by virtue of* representing the particular state or act de re. In all such cases, part of what makes it the case that the agent occupies that particular mental state, or performs that particular mental act, is, in part, her representing that state or act *de re*.

(b) The de re representation that is involved in constituting those state or acts doesn't occur until some time after the agent begins to occupy

> that state or perform that act. When the agent begins to occupy the state or perform the act, it is not yet determinate what state or act it is – that becomes determinate only once the agent represents the state or act in question de re, which is precisely what she does in response to questions of the form "do you believe that p?" or "why do you believe that p?"

Thus, cases that exhibit the puzzling phenomenon that we've described are analogous to the chess move described in the preceding section: you've lifted your queen and have thereby begun to move it either to f7 or g7, but you haven't yet decided which. The way you decide which of these two moves you are making is by considering the positions on the board. But by considering the positions on the board, you to decide to move your queen to g7, and thereby *make it the case* that your current move is that of moving your queen to g7. Once you've made it the case that your current move is that of moving your queen to g7, some version of the detectivist story can explain how you come to know that you've done so. But of course you cannot know that you've done so until you've done so, and (since the move is an essentially represented particular) you cannot have done so until you've represented your move de re as one of doing so. Thus, at the moment that you begin to move your queen, the move that you've begun to make could be a move to g7 or a move to f7, but it is not yet metaphysically determinate which of those two moves it is. This becomes determinate only when your decision (made by considering the positions of pieces on the board) makes it so, and only by virtue of your decision making it so. Of course, once you've decided which move to make, there is typically nothing puzzling about how you know that that is what you decided. Such knowledge does not exhibit the transparency phenomenon that we've been concerned to explain, and so the puzzle of transparency does not arise with respect to such knowledge. Once we've decided to move our queen to g7, we can simply report that we've decided this, and simply report that this is the move that we're making, without having to consider any further facts, extra-mental or otherwise.

We can finally return to the question I raised at the end of section I:

"Our puzzle was this: how can consideration of the reasons for or against p tell us what we believe? Moran's solution was to say that we are entitled to assume that whatever conclusion we draw from consideration of the reasons for or against p is a conclusion that we believe to be true. And it is plausible that we are entitled to assume this, because what we thereby assume is nothing more than what is implicit in our concept of *drawing a conclusion from some considerations*. But understanding what is implicit in that concept can help us to know what it is that we, on a particular occasion, believe *only if* we know what conclusion we draw, on that same occasion, from some considerations. But what explains how we know the latter? Let's grant that, in reviewing some considerations concerning some non-psychological

topic, we come to know that we are drawing a conclusion from those considerations; our question now is *how* we come to know this. For reasons that we will eventually discuss, Moran doesn't explicitly address that question – not in the quote above, nor in any other writing."

Why doesn't Moran address this question? I suggest that he doesn't address the question because he thinks – quite plausibly – that knowing what conclusion we have just drawn from some considerations is a form of self-knowledge that does not exhibit transparency. Once I've drawn a conclusion, I don't need to consult any extra-mental facts to realize that I've drawn this conclusion. And so, for all that Moran cares, some form of detectivism may provide a correct explanation concerning such knowledge.[9]

[9] I am grateful to Eli Chudnoff, Eric Marcus, John Phillips, John Schwenkler, and Sarah Wright for comments.

WORKS CITED

Armstrong, D., 1968/1993, *A Materialist Theory of the Mind*. London: Routledge.

Barnett, D., 2016. "Inferential Justification and the Transparency of Belief," *Nous* 50: 184 – 212.

Boghossian, P., 2012, "What is Inference?" *Philosophical Studies* 169: 1 – 18.

Boyle, M., 2009, "Two Kinds of Self-Knowledge," *Philosophy and Phenomenological Research* 78: 133-164.

---., 2011, "Transparent Self-Knowledge," *Proceedings of the Aristotelian Society Supplementary Volume* 85: 223 – 40.

Burge, T., 1996, "Our Entitlement to Self-Knowledge," *Proceedings of the Aristotelian Society* 96: 91-116.

Byrne, A., 2011, "Transparency, Belief, Intention," *Proceedings of the Aristotelian Society Supplementary Volume* 85: 201 – 21.

Evans, G., 1982, *The Varieties of Reference*, Oxford: Oxford University Press (ed. J. McDowell).

Gertler, B., 2011, *Self-Knowledge*. Routledge.Kind, A., 2003, "Shoemaker, Self-Blindness, and Moore's Paradox," *Philosophical Quarterly* 53: 39–48.

---., 2016, "Self-Knowledge and Rational Agency: A Defense of Empiricism," *Philosophy and Phenomenological Research* 95: 1 – 19.

Haugeland, J. 1982, "Weak Supervenience," *American Philosophical Quarterly* 19: 93 – 103.

Moran, R., 2003, "Responses to O'Brien and Shoemaker," *European Journal of Philosophy* 11: 402–19.

Neta, R., 2011. "The Nature and Reach of Privileged Access," in *Self-Knowledge*, Anthony Hatzimoysis, ed. Oxford: Oxford University Press.

Shoemaker, S., 1994, "Self-Knowledge and 'Inner Sense'," *Philosophy and Phenomenological Research* 54: 249–314.

Thomson, J., 1965, "Reasons and Reasoning," in *Philosophy in America*, Max Black, ed. Ithaca, N.Y.: Cornell University Press.