

ARTICLE

# The Role of Memory in Agential Self-Knowledge

Ben Sorgiovanni

The Queen's College, University of Oxford, Oxford, UK  
Email: [ben.sorgiovanni@queens.ox.ac.uk](mailto:ben.sorgiovanni@queens.ox.ac.uk)

## Abstract

Agentialism about self-knowledge (hereafter simply “agentialism”) is the view that key to understanding our capacity for self-knowledge is appreciating the connection between that capacity and our identities as rational agents—as creatures for whom believing, intending, desiring, and so on are manifestations of a capacity to be responsive to reasons. This connection, agentialists maintain, consists in the fact that coming to know our own minds involves an exercise of our rational capacities in the service of answering the relevant first-order question. Agentialists face the task of accounting for the connection between our identities as rational agents and our capacity to know our *stored* beliefs. It's plausible that one comes to know that one believes that *p* by exercising one's rational capacities in those cases where the belief that *p* is formed on the basis of present consideration of the reasons for and against *p*. But what exactly is the relevance of our rational capacities in the case where one has already formed the belief in question? In this paper I provide an answer to this question. That answer involves an appeal to a particular model of memory. According to the model I favor, memory preserves, in addition to the content of one's beliefs, one's commitment to their truth.

**Keywords:** Memory; agentialism; transparency; self-knowledge; rational capacity

## 1. Introduction

Agentialism about self-knowledge (hereafter simply “agentialism”) is the view that key to understanding our capacity for self-knowledge is appreciating the connection between that capacity and our identities as rational agents—as creatures for whom believing, intending, desiring, and so on are manifestations of a capacity to be responsive to reasons. This connection, agentialists maintain, consists in the fact that coming to know our own minds involves an exercise of our rational capacities in the service of answering the relevant first-order question.<sup>1</sup> (A rational capacity is a capacity which is governed by rational norms—norms relevant to the appraisal of an attitude or agent as rational or irrational. Inference is an example of a rational capacity; digestion is an example of a nonrational capacity.) Agentialists are typically impressed by the fact that often, I can answer the question whether I believe that *p* (for instance, whether I believe that it will rain tomorrow) simply by answering the question whether *p* (the question whether it will rain tomorrow). Let us say that when one answers the question whether one believes that *p* simply by answering the question whether *p*, one employs the *transparency procedure*. This procedure is significant, agentialists claim, because in the paradigmatic case, at least, it involves deploying one's rational capacities in the service of answering a question about what one believes: in the paradigmatic case, I answer the question whether I believe that *p* by *considering the reasons for and against p*.<sup>2</sup>

<sup>1</sup>Proponents of agentialism thus understood include Moran (2001, 2004, 2012) and Boyle (2009, 2011, 2017, 2019).

<sup>2</sup>Not all philosophers who emphasise the significance of the transparency procedure take it to reveal something important about the connection between our capacity to know our own minds and our identities as rational agents. See, for instance, Byrne (2011). Byrne's account emphasises our identities as rational, but not as rational *agents*.

It seems, however, that not every instance in which one employs the transparency procedure is one in which one must consider the reasons that bear on the relevant first-order question. Consider, for instance, one's stored beliefs—those beliefs which are in no way before one's mind (Frise 2017). I can come to know that I believe that  $p$ , say, where my belief that  $p$  is stored, by employing the transparency procedure—by answering the first-order question whether  $p$ . But it's not the case that my answering that question necessarily involves me considering the reasons for and against  $p$ . In cases where I have a stored belief that  $p$ , I can answer the question whether  $p$  simply by *remembering* that  $p$ . And I can remember that  $p$  without rehearsing the considerations which speak in favor of  $p$  or that initially convinced me that  $p$ . It's plausible that exercising one's capacity for deliberating on the first-order considerations is the way one comes to know that one believes that  $p$ , where the belief that  $p$  is formed on the basis of the present exercise of that very capacity—on the basis of present consideration of the reasons for and against  $p$ . But it's not at all clear that one must exercise that capacity in order to know that one believes that  $p$ , where the belief that  $p$  is stored. But then, if this capacity is not necessarily exercised in coming to know one's stored beliefs, what exactly is the relevance for self-knowledge of our rational capacities in these cases?

Call this “the stored belief question.” As long as the stored belief question goes unanswered, agentialism is radically incomplete. According to agentialists, key to understanding our capacity for self-knowledge is understanding the way in which self-knowledge involves an exercise of our rational capacities in the service of answering the relevant first-order question. But if it is not clear how the exercise of those capacities is involved in our coming to know our stored beliefs, then we have no reason to think that it *is* key to an account of how we know what we believe. After all, stored beliefs make up the vast majority of those of our beliefs which we are in a position to know at any given time.

In this paper, I provide an answer to the stored belief question. In section 2, I consider Richard Moran's response to the question as pressed by Nishi Shah and David Velleman. I suggest that Moran's response—specifically, his reminder that in order for an utterance to count as the expression of a belief of mine, it must bear *some* relation to the exercise of my rational capacities for determining what is true and false—should reassure the agentialist. But Moran's response raises a question: How could my utterance “ $p$ ” bear a relation of the appropriate sort to my rational capacities for determining what is true and false if it's not the upshot of deliberation in the present moment about whether  $p$ ? In sections 3 and 4, I argue that it can be in virtue of my remembering that  $p$  that my utterance bears this relation. I sketch two models of remembering in the semantic mode: the judgment model and the testimony model. I go on to argue that we have good reason to prefer the judgment model. Unlike the testimony model, it characterizes instances of remembering in a way which is not at odds with rational norms governing change in the rational bases of our beliefs over time. In section 5, I argue that this reason for preferring the judgment model grounds an answer to the stored belief question.

## 2. The stored belief question: Moran's response to Shah and Velleman

Nishi Shah and David Velleman are among those philosophers who have pressed the question of stored beliefs against agentialism. They begin by pointing out that the question “Do I believe that  $p$ ?” is ambiguous. It can mean either “Do I already believe that  $p$  (that is, antecedently to considering this question)?” or “Do I now believe that  $p$  (that is, now that I am answering the question)?” (2005, 506). One is in a position to answer the latter question precisely by forming the belief that  $p$ . But not so, according to Shah and Velleman, in the case of the former question: “[O]ne cannot answer the question *whether I already believe that p* in a way that begins with forming the belief” (506; emphasis in original). Shah and Velleman go on to acknowledge that asking whether  $p$  can indeed be a means of determining whether one already believes that  $p$ . But it's important, they claim, that one refrains from any reasoning as to whether  $p$ :

If the question is *whether I already believe that p*, one can assay the relevant state of mind by posing the question *whether p* and seeing what one is spontaneously inclined to answer.

In this procedure, the question *whether p* serves as a stimulus applied to oneself for the empirical purpose of eliciting a response ... But the procedure requires one to refrain from any reasoning as to whether *p*, since that reasoning might alter the state of mind that one is trying to assay. Hence, asking oneself *whether p* must be a brute stimulus in this case rather than an invitation to reasoning. (506; emphasis in original)

Compare the case in which I'm seeking to answer the question whether I now believe that *p*. I can settle this question by reasoning to a conscious judgment whether *p*, because "the question whether I now believe that *p* naturally leads to the question whether to judge that *p*, which gives way to the question whether *p*, now posed as an invitation to reasoning" (507).

I take it that by "reasoning," Shah and Velleman do not mean *explicit reasoning*—the sort of reasoning involved in deliberation, for example. Instead, "reasoning" should be understood broadly to encompass any case in which one makes up one's mind that *p* on the basis of reasons for and against *p*. One does not need to explicitly consider those reasons. (Indeed, proponents of agentialism are clear that the rational capacity they are interested in is not exhausted by our capacity for deliberation or explicit reasoning. See, for instance, Moran [2004, 458].) If Shah and Velleman are right, then reasoning—whether implicit or explicit—*cannot* be a way of coming to know whether one believes that *p*, where the belief that *p* is stored, for such reasoning runs the risk of altering the target belief.

Richard Moran—a prominent proponent of agentialism—has defended the view against Shah and Velleman's line of objection in the following way. Not just any response to a question about what I believe will count as an expression of my belief, Moran notes, even in the case where the belief in question is stored. If someone asks me, "What do you believe is the capital of Spain?" and I utter, "Madrid," my utterance will count as an expression of my belief that Madrid is the capital of Spain only if certain conditions are met. Paramount among these is the requirement that I hold it *true* that Madrid is the capital of Spain. But, Moran goes on:

[I]f I relate to my "stored" belief as something I take to be true, it will be hard to see how I can see my relation to it, however spontaneous, as insulated from the engagement of my rational capacities for determining what is true or false. It cannot, for instance, be seen by me (or my auditors, if they are the ones applying the stimulus) as simply some name that is produced upon receiving the stimulus, for it has to represent what I take to be true, as an answer to the question asked. Hence, there is considerable background involving my rational agency that has to be assumed for my response to the stimulus to count as my spontaneous answer to the question. (2012, 221)

Moran's point is surely right: in order for an utterance to count as the expression of a belief of mine, it must bear *some* relation to the exercise of my rational capacities for determining what is true and false. After all, my utterance will only count as the expression of my belief if the proposition it expresses is one which I hold to be true. But by itself this point is insufficient to satisfactorily answer the question of stored beliefs. To satisfactorily answer that question, the agentialist needs to offer some account of what exactly the relation *is* between the exercise of my rational capacities and my ability to know that I believe that *p*, where the belief that *p* is stored; it is not enough simply to note that some such relation must obtain.<sup>3</sup>

Now, Moran surely realizes this. Indeed, certain things which he goes on to say suggest a broader point—namely, that even in the case of stored beliefs, one comes to know that one believes that *p* by reflecting on the reasons for and against *p*:

For any mental content, word association, or exclamation that may be produced by the stimulus, it will count as relevant to the question of my belief only if I am relating to that

<sup>3</sup>See Cassam (2014, 114ff.) on this point.

content or word as representing my belief about that matter. *That, however, will mean engaging with my rational capacities in a way that involves my reflecting on the facts of the matter*, which was supposed to be excluded from the process because it was thought to “contaminate” it. But if that were so, then it is hard to see how a person could ever perform the ordinary task of telling us what they already believe about something. (2012, 222; emphasis added)

Again, Moran is surely right that I must relate appropriately to the content or utterance which is produced by a stimulus in order for that content or utterance to bear on the question of what I believe. But it is questionable whether it somehow follows from this that in order for an utterance, say, to bear on this question, it must be the upshot of my reflecting on the “facts of the matter,” by which Moran presumably means those facts which bear on the relevant first-order question. Leaving aside Shah and Velleman’s concerns about the risk of “contamination” which such reflection allegedly introduces, it just seems empirically false that we know what we think about some issue only by reflecting in the present moment on the relevant first-order considerations. Suppose I’m asked, “What do you believe is the capital of Spain?” and I respond, “Madrid.” It’s true that in order for this response to count as the expression of my belief, I must hold it true that Madrid is the capital of Spain. But it’s not necessary that I first consider the reasons for and against Madrid’s being the capital of Spain. I can simply *recall* that Madrid is the capital of Spain.

In fairness to Moran, he acknowledges that the sort of rational activity he has in mind need not be explicit (2012, 222). But my point is that there need be no consideration of the reasons for and against *at all*. Indeed, one may have forgotten the considerations that initially convinced one—and which might now convince one—that, for example, Madrid is the capital of Spain. Still, one’s utterance may count as the expression of one’s belief.<sup>4</sup>

For these reasons, I don’t think Moran’s suggestion—the suggestion that in order for an utterance to count as the expression of a belief it must be the result of present (implicit or explicit) reflection on the first-order question—can be quite right. Still, I think that what *is* right about Moran’s response should reassure the agentalist. What is right about Moran’s response is that in order for an utterance to count as the expression of a belief of mine, it must bear some relation to the exercise of my rational capacities for determining what is true and false. But now we want to know: how *could* my utterance “*p*” bear a relation of the appropriate sort to my rational capacities for determining what is true and false, if it’s *not* the upshot of deliberation in the present about whether *p*? In sections 3, 4, and 5, I’ll offer an answer to this question. My answer will appeal to the functioning of memory. I want to say that it can be in virtue of your remembering that *p* that your utterance bears the relation in question. I begin by detailing the sort of memory that I’m interested in.

### 3. Models of remembering in the semantic mode

#### 3.a The episodic/semantic distinction

“Memory,” in the most general sense, is that capacity, whatever it is, which allows for the retrieval of information learned at some earlier time. “Memories” are those contentful mental items which are retrieved *in instances of remembering*. I distinguish, then, between memory as a capacity, token memories, and acts of remembering.

There are various ways of discriminating memory capacities.<sup>5</sup> One popular taxonomy is hierarchical in structure. At the top of the hierarchy is the distinction between “declarative” and “nondeclarative” memory. The distinction is meant to mark the difference between those capacities

<sup>4</sup>I discuss Moran’s response to the stored belief question, and the differences between his response and my own, in section 4.

<sup>5</sup>See Werning and Cheng (2017) for an overview.

which allow, respectively, for the retrieval of memories a subject can express (for example, remembering that your rental inspection is tomorrow) and for the retrieval of memories which the subject can only demonstrate (for example, remembering how to ride a bike) (Bernecker 2009, 12; Squire and Zola-Morgan, 1988; Squire 2009).

Following Tulving (1972), it's standard to distinguish, within the category of declarative memory, between "episodic" and "semantic memory." The question of how exactly to draw the distinction between these two memory capacities is contested. One popular proposal is that remembering in the episodic mode has a distinctive phenomenology; specifically, in instances of episodic recall, past events are presented by way of past experiences, which are undergone at the moment of recall.<sup>6</sup> Let us say that episodic memory is that capacity, whatever it is, which allows one to recall information by re-experiencing certain of one's past experiences (or, more exactly, experiences which are presented to one as one's past experiences<sup>7</sup>). Let us say that semantic memory is that capacity, whatever it is, which allows us to recall information *without* reexperiencing such experiences. If, for instance, I remember that I once visited a beach in New Zealand by reexperiencing the feeling of the sand between my toes, then the capacity involved is episodic memory. If, however, I do not reexperience this feeling or any other past experience, then the capacity involved is semantic memory.<sup>8</sup> (Of course, in most cases the exercise of semantic memory results in one's calling to mind, not information about one's past, but generic information about the world—for example, that geese can be territorial, that exercise is good for you, and so on and so forth.)

We can extend the semantic/episodic distinction to both memories and acts of remembering. Let us say that a memory is semantic if it is poised for retrieval through the exercise of semantic memory. And let us say that one remembers that *p* in the semantic mode if one remembers that *p* without reexperiencing certain of one's past experiences. Episodic memories and remembering in the episodic mode can be characterized similarly.

My concern in this paper is first and foremost with remembering in the semantic mode—and, to a lesser extent, with semantic memory and semantic memories. I will have nothing to say about their episodic analogues. This is not because there is any privileged connection between recalling our stored beliefs and remembering in the semantic mode. Plausibly, we recall many of our stored beliefs in the episodic mode—by reliving certain experiences. My reasons for focusing on semantic memory have, rather, to do with issues of manageability. The nature of episodic memory is hotly disputed. Philosophers disagree, for instance, about how to understand the content of episodic memories.<sup>9</sup> I wish to steer a path around these debates. Some of what I say will, I expect, apply with equal force to episodic memory, some may not.

### 3.b Models of remembering in the semantic mode: Testimony vs. judgment

There are at least two competing models of remembering in the semantic mode, which I call the "testimony" and "judgment" models, respectively.<sup>10</sup> According to the testimony model, remembering that *p* involves calling to mind that *p*. But this calling to mind does not amount to believing

<sup>6</sup>See, for example, Rowlands (2009, 318).

<sup>7</sup>This qualification is important so that we do not rule out the possibility of quasimemory by definition. See Parfit (1984). For an argument against the notion of quasimemory, see Schechtman (1990).

<sup>8</sup>Some philosophers have suggested that even this drawing of the distinction is not fit for purpose. See, for instance, Bernecker (2009, 13ff.). Bernecker prefers what he calls a grammatical taxonomy—a classification in terms of memory's grammatical objects. Nothing about my argument depends on our using the semantic/episodic distinction as opposed to Bernecker's classificatory scheme.

<sup>9</sup>For competing views, see Searle (1983, chap. 3), Fernandez (2006) and Recanati (2007, chap. 18).

<sup>10</sup>I introduce both models against the backdrop of a broadly representational theory of belief—a theory according to which believing that *p* consists in being appropriately related to a representation that *p*. (Neither assumes a particular view about what representations are.) But it seems to me one might endorse similarly spirited models even if one rejects the representational theory of belief.

that *p*. One must first endorse the content one has called to mind as true. On this sort of view, remembering is, in an important sense, like receiving testimony: in both cases content is presented to one as true, and in both cases one may endorse or reject (or adopt some more nuanced epistemic attitude towards) what is presented to one. Even in the case in which one does endorse it, the act of endorsing it is distinct from the act of remembering, just as being told that *p* is distinct from endorsing that *p* on that basis.<sup>11</sup> According to the testimony model, remembering that *p* in the semantic mode never amounts to believing that *p*.

According to the judgment model, what occurs when one remembers that *p* does amount to believing that *p*. Remembering is not like receiving testimony. Rather, it is akin to judgment: just as judging that *p* is not accompanied by some further act—the act of endorsing that judgment—nor is remembering that *p* accompanied by some further act—the act of endorsing the deliverances of memory as true. To judge that *p* just is to endorse that *p*. So too, to remember that *p* just is to endorse that *p*. Note that a proponent of the judgment model does not want to say that to remember that *p* just is to judge that *p*. Indeed, the act of remembering that *p* does not even involve the act of judging that *p*, on her view. The claim is simply that consciously judging that *p* provides a useful model on which to think of remembering that *p*.<sup>12</sup>

I devote the remainder of this subsection to clarifying the testimony and judgment models, respectively.

Proponents of the testimony model think of remembering that *p* in the semantic mode on the model of receiving testimony that *p*. However, a proponent of the testimony model can (and should) accept that there are important differences between memory and testimony. One apparent difference is that remembering that *p* is factive, while being told that *p* is not. One can testify that *p* whether *p* is true or false. But, plausibly, one can remember that *p* only if *p* is true.<sup>13</sup> For this reason, it isn't quite right to say that remembering that *p* is like being told that *p*, on the testimony model. It is more exact to say that remembering that *p* is like being informed or made aware that *p*.

Proponents of the judgment model, too, will want to qualify their view, in their case by acknowledging apparent differences between remembering that *p* and consciously judging that *p*, differences which I won't do more than merely list here. First, memory is factive while judgment clearly is not. (I can judge that *p* even when *p* is false.) Second, plausibly, one counts as remembering that *p* only if the content's coming before one's mind is appropriately causally connected to the occasion of one's learning that *p*.<sup>14</sup> There does not seem to be any analogous condition in the case of judgment. A third apparent difference concerns the way in which the contents of memory and judgment are, respectively, presented to one. In paradigmatic instances of remembering, one might think the relevant propositional content is presented as content the truth of which one has a standing commitment to. Not so, one might think, in the case of judgment. Clearly, each of these points stands in need of elaboration. But suffice it to say that they are points which a proponent of the judgment model can happily accept. In spite of these differences, she can insist, judging that *p* provides a useful model on which to think of remembering that *p* in the semantic mode.

<sup>11</sup>Some of the things which Dummett (1993) says suggest this view.

<sup>12</sup>For an explicit defence of this model, see Owens (2000, chap. 10). See also Barnett (2015); Note that the testimony and judgment models are first and foremost views concerning the structure of, as opposed to the warrant(s) involved in, remembering that *p*. The models may conduce towards particular views about the nature of our warrant for memorial beliefs. But it seems that one can endorse either model without thereby committing oneself to any particular view. For example, one can endorse the testimony model and yet think that one has a standing entitlement to believe that one's memory is a reliable guide to the truth. Similarly, one can endorse the testimony model and consistently claim that the relevant warrant is a posteriori—a warrant which depends on one's having evidence that, for instance, one's memory is functioning reliably, that the target proposition is in fact true, and so on. The analogous point holds true for proponents of the judgment model.

<sup>13</sup>Although, see Hazlett (2010).

<sup>14</sup>For discussion of what it is for this causal connection to be appropriate, see Bernecker (2009, chap. 5).

Indeed, there is a further qualification which a proponent of the judgment model ought to make to her view. Many philosophers are convinced that one might remember that  $p$  and yet, at the time of remembering, fail to believe that  $p$ . (See, for instance, Bernecker [2009] and Martin and Deutscher [1966].) When asked which is the largest ocean in the world, she may correctly recall that it is the Pacific Ocean. But she may doubt the reliability of her memory to such an extent that in spite of her recalling the answer correctly, she does not believe that the largest ocean in the world is the Pacific Ocean. Still, it seems right to say that she has remembered that it is. If such cases are indeed genuine cases of remembering, then they seem to pose a problem for the judgment model.<sup>15</sup>

In my view, such cases give us good reason to think that believing that  $p$  is not necessary for remembering that  $p$ . But this does not pose a problem for the judgment model as long as we understand it as a model of *paradigmatic instances* of remembering that  $p$ . Thus understood, a proponent of the judgment model is someone who thinks that paradigmatic cases of remembering that  $p$  involve believing that  $p$ . This qualification allows that there might be nonparadigmatic cases—cases of remembering which don't involve believing that  $p$ . Such cases might nevertheless count as instances of remembering insofar as they resemble paradigmatic cases in certain key respects (for example, the content remembered is appropriately causally related to the occasion of one's having learned that  $p$ ). Of course, this qualification will be unsatisfactory if it is a condition of adequacy for an account of memory that it provides a necessary condition (or a set of such conditions) for remembering. But we should only accept that this is a condition of adequacy if we have reason to think that there is one thing (or set of things) common to everything which we want to call remembering in the semantic mode. And it's not obvious that this is the case. Consequently, it's not obvious that we should favor an analysis of memory over an account that seeks to offer a characterization of paradigmatic (and/or nonparadigmatic) instances of remembering. (Hereafter, when I talk about the judgment model, I take for granted the qualification outlined in this paragraph.)

#### 4. Remembering in the semantic mode and motivating epistemic reasons

Both the testimony and judgment models of remembering in the semantic mode ground an explanation of how my uttering " $p$ " in response to the question "Do you believe that  $p$ ?" might count as an expression of my belief—an expression of my commitment to the truth of  $p$ —without that utterance being the upshot of present implicit or explicit reasoning about the question whether  $p$ . In the case of the judgment model, the explanation is simply that, in response to the question, I remember that  $p$ . Since to remember that  $p$  is just to believe that  $p$ , we have all we need to explain how my uttering " $p$ " could count as an expression of my belief. In the case of the testimony model, the explanation is that my utterance is the upshot of my accepting the deliverances of my memory. Unlike the judgment model, the explanation involves an appeal to two distinct acts: the act of remembering and the act of affirming or accepting that which one remembers.

Are there grounds for preferring one model to the other? In this section, I argue that there is good reason to prefer the judgment model. The reason has to do with the rational bases for belief and the norms which govern changes in those rational bases over time.

On closer inspection, the picture of remembering in the semantic mode which the testimony model supplies may well strike us as odd. It seems to be a picture according to which we have to "remake" our beliefs at each successive moment of remembering. The model seems to exclude outright the possibility that semantic memory might preserve our beliefs across time.<sup>16</sup> And yet we

<sup>15</sup>See Debus (2010) for an argument as to why such cases are not correctly described as cases of remembering.

<sup>16</sup>One might wonder whether this is really so. Perhaps a proponent of the testimony model could insist that to have a standing belief that  $p$  is simply to be disposed to assent to the proposition that  $p$  whenever one remembers it. If this is a plausible account of standing beliefs, then when I affirm that  $p$  on the basis of having remembered that  $p$ , I am *not* "remaking" my belief that  $p$ . Rather, I'm manifesting the disposition which is constitutive of my having a standing belief that  $p$ . In fact, I don't think that this

might well think that one of the principal functions of memory is to allow us to rely on past judgments in order to free us from the responsibility of having to repeat those judgments in the present. Of course, we are always at liberty to reaffirm, adjust, or relinquish a stored belief at the moment of recall, but the picture supplied by the testimony model seems to make this a *requirement* of remembering. The concern is that the picture does not do justice to the sense in which remembering is, intuitively, *recalling something to which one is committed*.

Now stated so plainly, it is open to the proponent of the testimony model simply to deny the intuition. What is needed is some argument for thinking that remembering in the semantic mode can consist in recalling something to which one is committed. I think that there is such an argument. It involves *motivating epistemic reasons*, where a motivating epistemic reason for believing that *p* is a reason which convinces one that it is true that *p*—a reason that underwrites a rational explanation of why one believes that *p* (Barnett 2015; Turri 2009). (Hereafter, I refer to motivating epistemic reasons simply as “motivating reasons.”) Four clarifications are in order. First, motivating reasons are not necessarily *good* reasons—reasons which count in favor of, justify, or make reasonable the target belief. One may be convinced that *p* on the basis of considerations which are in fact bad reasons for believing that *p*. As long as the considerations convince one, they are among one’s motivating reasons. Second, not every reason which enters into an explanation of why one believes that *p* is among one’s motivating reasons. Perhaps the play of light on a hot day is among *the reasons why* I believe that there is water on the road ahead. But it is not among *my reasons for* believing—it is not a consideration which *convinces* me that there is water on the road ahead. Third, I take no particular view about what motivating reasons are. I leave it open whether they are, for instance, propositions (Scanlon 1999), nonmental facts (Dancy 2000), or mental states (Turri 2009). Fourth, it is consistent with the way I am thinking of them that one might have multiple motivating reasons for a target belief.

The argument in support of the intuition that semantic memory can consist in recalling something to which one is committed involves cases like the following. Suppose that my motivating reason for believing that the speed of sound is 343 m/s is my physics teacher’s telling me that it is; my teacher’s testimony is the consideration that convinces me that the speed of sound is 343 m/s. Suppose that at some time subsequent to my having formed the belief, I remember that the speed of sound is 343 m/s, but do not call to mind my reasons for thinking that it is. I haven’t *forgotten* my teacher’s testimony, I simply do not bring it to mind. Finally, suppose that in the time between my first learning that the speed of sound is 343 m/s and my recalling it, I neither forget nor learn anything. I take it that the natural thing to say about the case as described is that my motivating reason for believing that the speed of sound is 343 m/s is, at the moment of recalling this information, my teacher’s testimony.

And yet, it seems that a proponent of the testimony model *cannot* say this. According to the testimony model, my recalling that the speed of sound is 343 m/s involves calling to mind the proposition that the speed of sound is 343 m/s, where my calling it to mind does not amount to my believing that the speed of sound is 343 m/s. It is only once I endorse the content that I have called to

---

account of standing beliefs is plausible. For the account to succeed, one’s remembering that *p* must be conceived in such a way that it does not presuppose one’s having a standing belief that *p*, otherwise the account would be circular. Remembering that *p* cannot be calling to mind the belief that *p*; it must instead be something like calling to mind the proposition that *p*.

But now, suppose that I have been considering the proposition that *p*, but am unsure whether to affirm or deny it. Suppose that I now learn some new information from which it follows obviously that *p*, but that I do not return to question whether *p*. If I did, I would affirm that *p* without hesitation. I seem to have the requisite disposition for a standing belief that *p*—the disposition to assent to *p* upon remembering it. But do I have a standing belief that *p*? Plausibly not. The inference from my new information to that *p* may be obvious, but I have not drawn it. Moreover, it may be that I continue to act in a way which suggests agnosticism about whether *p*. If I do believe that *p*, it does not play the role in regulating my thought and action which a belief that *p* normally does. (See Audi [1994] for this line of argument against the model of belief under consideration.)

I thank an anonymous referee for pressing me to say more on this issue.



mind that I form the belief. In cases like the one at hand—cases in which one endorses the content without first calling to mind considerations which earlier convinced one—the consideration which presently convinces one, it seems, is just the calling to mind of the content (coupled, perhaps, with beliefs about the reliability of one’s memory, one’s general consciousnesses as an epistemic agent, and so on and so forth). In the case at hand, my teacher’s testimony, though not forgotten, seems irrelevant to a rational explanation of why it is that I now believe that the speed of sound is 343 m/s.<sup>17</sup> Unlike the calling to mind of the content, my teacher’s testimony seems to play no role in convincing me that the speed of sound is 343 m/s.<sup>18</sup>

This is not the case on the judgment model. On the judgment model, memory preserves belief. When I remember that the speed of sound is 343 m/s, I call to mind my *belief* that the speed of sound is 343 m/s. The question whether the speed of sound is 343 m/s is not open for me at the moment of recall in the way that it is on the testimony model. Consequently, in cases like the case at hand—in which one does not call to mind those considerations which earlier convinced one—there is no pressure to identify considerations presently before one’s mind as those which rationally explain one’s belief at the moment of recall. Those considerations which rationally explain one’s belief can be precisely those considerations which initially convinced one.

Indeed, it seems that we can *only* make room for this possibility if the judgment model is correct. Any account according to which remembering that *p* consists in calling to mind something which falls short of belief will have to offer some account of the rational basis for your believing that *p* on the basis of recalling that *p* in cases like the one at hand—cases in which you recall that *p* without recalling the considerations which initially convinced you that *p*. And it is difficult to see how any such account could avoid the conclusion to which the proponent of the testimony model seems committed—namely, that the rational bases for your belief that *p* are the deliverances of memory, not those considerations which initially persuaded you. To this extent, we have reason to prefer the judgment model to the testimony model.

Unlike the judgment model, the testimony model seems committed to describing cases like the one at hand in a way which is deeply counterintuitive. Specifically, the testimony model seems committed to the view that, at the moment of recall, one’s motivating reason for believing that *p* differs from one’s original motivating reason for believing that *p*. But intuitively, one’s motivating reason at the moment of recall for believing that *p* is just the consideration(s) which initially convinced one that *p*.

A proponent of the testimony model might challenge the claim that this result is counterintuitive. It cannot be counterintuitive, they might point out, simply in virtue of the fact that it amounts to supposing that one’s motivating reasons may change over time. After all, the rational bases for our beliefs often *do* change over time, even in cases where we have not forgotten the evidence that originally convinced us. But then, what is so objectionable about the notion that one’s calling content to mind might take the place of one’s original motivating reason in the sorts of cases we’re considering?

Clearly, the rational bases for one’s beliefs can change, even in cases where one has not forgotten the considerations that originally persuaded one. But insofar as one is rational, such change will conform to certain norms. We are not simply creatures who respond to reasons at a particular time; our responses have duration. We do not simply form beliefs; we maintain them. And many of those rational norms which govern the formation of beliefs at a particular time also govern their maintenance. Here is an example of such a norm:

(*Strongest Evidence*): One’s (primary) motivating reason(s) for believing that *p* ought to be what one reasonably takes to be one’s strongest evidence that *p*.

<sup>17</sup>Of course, my teacher’s testimony is not irrelevant to a purely causal explanation of why it is that I now believe that the speed of sound is 343 m/s. After all, if I hadn’t received that testimony, I wouldn’t have formed the belief. And if I hadn’t formed the belief, I wouldn’t now be in a position to call to mind the proposition that the speed of sound is 343 m/s.

<sup>18</sup>Barnett (2015, 379, 388) reaches similar conclusions about similar cases.

(Strongest Evidence), or something like it, governs belief formation. But plausibly, it also has implications for the way in which we ought to update our motivating reason(s) for believing that  $p$  where the belief that  $p$  is stored. According with the norm, it seems, requires that one update one's motivating reason(s) only under certain conditions—specifically, only if either: (i) one is persuaded reasonably that certain (but not all) of the considerations which initially convinced one are not, in fact, sufficient evidence for believing that  $p$ ;<sup>19</sup> or (ii) one encounters what one reasonably takes to be better evidence that  $p$ .

In cases where (i) is satisfied, it is not that one ought to acquire a *new* motivating reason for believing that  $p$ , one ought simply to cease to be motivated by those considerations which one has reasonably lost confidence in. Initially, my smelling wood smoke and hearing fire trucks are the considerations which convince me that there is a fire in the valley. I come to learn that the wood smoke is emanating from my neighbor's barbecue. If I'm to accord with (Strongest Evidence), my motivating reason for believing that there is a fire in the valley now ought simply to be my hearing fire trucks. In cases in which (ii) is satisfied, one does necessarily acquire a new motivating reason. Initially, my hearing fire trucks is the consideration which convinces me that there is a fire in the valley. Upon approaching the valley, I see plumes of smoke rising above the treeline. Again, if I'm to accord with (Strongest Evidence), my motivating reasons for believing that there is a fire in the valley ought to include the plumes of smoke.

In the case involving my belief that the speed of sound is 343 m/s, neither (i) nor (ii) is satisfied. Indeed, in the case as described, it is trivially true that (i) is unsatisfied, for I have only one motivating reason for the target belief. But even if we were to adjust the case so that this consequence was avoided, (i) would remain unsatisfied. The change in the rational basis of my belief is not prompted by my being persuaded that my physics teacher's testimony is not sufficient evidence for believing that the speed of sound is 343 m/s. Nor is it a consequence of my encountering what I reasonably take to be better evidence for believing that the speed of sound is 343 m/s. In the case as described, I do not take my calling to mind the content that the speed of sound is 343 m/s to be better evidence than my teacher's testimony for believing that the speed of sound is 343 m/s. But even if I were to think it better evidence, I would not be *reasonable* in doing so. Arguably, there are cases in which the mere fact that I remember that  $p$  constitutes *some* evidence that  $p$ . (I might remember that Madrid is the capital of Spain but have forgotten all my original evidence for thinking that it is. Still, If I'm justified in believing that my memory is generally reliable, that I'm a responsible epistemic agent, and so on, the mere fact that I remember it might constitute some evidence for Madrid's being the capital of Spain.) But in cases where I still have my original evidence available, my remembering that  $p$  cannot constitute *better* evidence that  $p$ . If it did, I could improve my epistemic situation simply by remembering what I have learned. But this is absurd. Consequently, (ii) is not satisfied either.

We have, then, something to say in response to the proponent of the testimony model who presses us to say more about why exactly it is counterintuitive to suppose that in cases like the one at hand, one's motivating reason at the moment of recall is one's calling to mind the relevant content, not those considerations which initially convinced one. Such change in the rational bases for one's beliefs violates a norm which governs change in the rational bases of one's beliefs over time—namely, (Strongest Evidence).

In [section 2](#), I considered Moran's response to the stored belief question. I suggested that contrary to what Moran seems to suggest, it's not the case that in order for an utterance to count as the expression of a belief of mine, the utterance must be the consequence of present reflection on the reasons for and against the relevant first-order question. And yet it may now seem unclear how my

---

<sup>19</sup>If one is persuaded reasonably that all of the considerations which initially convinced one that  $p$  are not in fact sufficient evidence for believing that  $p$ , then what one ought to do is not, properly speaking, update one's motivating reasons for believing that  $p$ , one ought to cease believing that  $p$ .

own proposed response to the stored belief question avoids this conclusion.<sup>20</sup> I have defended the judgment model of memory on the grounds that it respects a norm which governs the updating of one's motivating reasons for believing that  $p$ —namely, (Strongest Evidence). But it might seem that according with this norm itself requires reflection on the first-order evidence, perhaps on the basis of something like the following chain of reasoning. (Strongest Evidence) requires updating one's motivating reasons for believing that  $p$  only if one's evidence that  $p$  has changed—specifically, only if one now has in hand stronger evidence that  $p$  than is constituted by one's existing motivating reasons. But in order to know whether one has such evidence in hand one must reflect, however cursorily, on one's evidence. But then, conforming with (Strongest Evidence) requires reflection on the evidence after all, for conforming with this norm requires knowing whether one's evidence has changed in a way which warrants updating one's motivating reasons.

In fact, this chain of reasoning rests on a subtle misreading of (Strongest Evidence). I also think it rests on a faulty premise. Let me first identify the faulty premise and then explain why it rests on a misreading.

It's not true that I can know at the moment of recall whether I have acquired stronger evidence that  $p$  only if I reflect at the moment of recall on the evidence that  $p$ . Earlier, I discussed a case in which my motivating reason for believing that the speed of sound is 343 m/s is my physics teacher's telling me that it is. Later, I remember that the speed of sound is 343 m/s. In order for this belief to be justifiably based at the moment of recall and on my teacher's testimony, do I first need to review all available evidence that the speed of sound is 343 m/s? No. I am entitled to rely on memory. Specifically, I am entitled to rely on memory's presenting my teacher's testimony as *the* reason to believe that the speed of sound is 343 m/s. I don't first need to *check* that I do not have any stronger evidence at hand for its being so. Granted, in the case where there is available higher-order evidence—evidence that there is, indeed, stronger available evidence that the speed of sound is 343 m/s—I'll need to do some reflection on the first-order evidence. But in the absence of such evidence, I am entitled to rely on the deliverances of my memory. Let us assume that it is *true* that at the moment of recall my strongest evidence for my belief is my teacher's testimony. In this case, I know that this is my strongest evidence—crucially, I know this without having to reflect on the first-order evidence.

I said that the chain of reasoning sketched above rests on a subtle misreading of (Strongest Evidence). (Strongest Evidence) does not require that one's motivating reasons for believing that  $p$  be one's strongest evidence to hand that  $p$ —simply that one's motivating reasons be those considerations which one *reasonably takes* to be one's strongest evidence that  $p$ . It may be that my motivating reason for believing that  $p$  is that  $q$ . Suppose that I also believe that  $r$ . Further, suppose that  $r$  is *in fact* stronger evidence than  $q$  that  $p$ . Still, if I reasonably believe that  $q$  is stronger evidence for  $p$ , then I am not violating (Strongest Evidence) in believing that  $p$  on the basis that  $q$ .

Does conforming to (Strongest Evidence) thus understood require reflection on the evidence? Do I need to reflect on the evidence in order to determine what I *reasonably take* to be my strongest evidence? Again, not in the sort of case I considered earlier. In that sort of case, I am in a position to “determine” what I take to be my strongest evidence simply by recalling my belief along with the associated motivating reason.<sup>21</sup> I do not need to consider all the available evidence that the speed of sound is 343 m/s. Do I need to reflect on the evidence in order to determine that I am being *reasonable* in taking my teacher's testimony to be my strongest evidence? I don't think so. Again, in the absence of evidence to the contrary, I am entitled to assume that it is my strongest evidence if my

<sup>20</sup>I thank an anonymous referee for pressing me to say more on this point.

<sup>21</sup>This is not, properly speaking, a way of *determining* what I take to be my strongest evidence for the target belief, for recalling that belief along with the associated motivating reason is not *evidence* that I take that reason to be the strongest evidence for the belief. Rather, my recalling it as my motivating reason *just is*, in the standard case at least, my taking it to be my strongest evidence.

memory presents it as such. In such a case, my judgment that I am being reasonable in taking this to be my strongest evidence is itself reasonable in the absence of reflection on the evidence. Given that I am indeed being reasonable in taking it to be my strongest evidence, I can *know* that I reasonably take it to be my strongest evidence—again, crucially, I can know this in the absence of reflection on the first-order evidence.

There remains, then, a crucial difference between my response to the stored belief question and the response sketched by Moran.

## 5. Conclusion

I began this paper by raising a question for agentialists—those who think that it is key to self-knowledge that it involves an exercise of our rational capacities in the service of answering the relevant first-order question. I called this the “stored belief question”: if one’s capacity to deliberate on the relevant first-order considerations is not necessarily exercised in coming to know one’s stored beliefs, what exactly is the relevance for self-knowledge of our rational capacities in cases where the target beliefs are stored? In section 2, I suggested that Moran’s response to Shah and Velleman should reassure the agentialist: in order for an utterance to count as the expression of a belief of mine, it must bear *some* relation to the exercise of my rational capacities for determining what is true and false. But it’s not the case—contrary to what Moran seems to suggest—that in order to bear this relation, the utterance must be the consequence of present reflection on the reasons for and against the relevant first-order question.

How could my utterance “*p*” bear a relation of the appropriate sort to my rational capacities for determining what is true and false if it’s not the upshot of deliberation in the present moment about whether *p*? In sections 3 and 4, I argued that it can be in virtue of my remembering that *p* that my utterance bears this relation. I sketched two models of remembering in the semantic mode—the judgment model and the testimony model—both of which are in a position to explain how it is that my response “*p*” could count as the expression of my belief. I went on to argue that we have good reason to prefer the judgment model. Unlike the testimony model, it respects constraints that govern the updating of motivating reasons for believing that *p*. These constraints follow from (Strongest Evidence)—the norm that one’s (primary) motivating reason(s) for believing that *p* ought to be what one reasonably takes to be one’s strongest evidence that *p*. (Strongest Evidence) has implications not only for those capacities involved in the formation of a belief at a time, its remit extends to those capacities involved in the preservation of a belief, together with its motivating reasons, through time.

But (Strongest Evidence) is a norm of rationality, so memory itself is a capacity which is governed by rational norms—and, hence, a rational capacity. We are now in a position to give the following answer to the stored belief question. Suppose that I am asked whether I believe that *p* in a case where my belief that *p* is stored. I am in a position to answer this question by employing the transparency procedure—by answering the first-order question whether *p*. But I answer this question not by forming a belief on the basis of consideration of the reasons for and against *p*, but by *recalling* that *p*. But the capacity which facilitates the recall of a belief formed at some earlier time—namely, semantic memory—is itself a rational capacity, for it is a capacity governed by rational norms. So the guiding agentialist thought is vindicated: even in the case of stored beliefs, I answer the question whether I believe that *p* by exercising my rational capacities in the service of answering the first-order question whether *p*. I do not, contra Moran, necessarily engage those capacities which agentialists emphasise in the case where I form the belief in the present moment—namely, those capacities involved in deliberation on the reasons for and against *p*. Nevertheless, my answering the question necessarily involves the exercise of certain of my rational capacities. The capacities involved are those which facilitate the preservation of our reason-responsive doxastic states, together with their motivating reasons, through time.

**Acknowledgments.** I am incredibly grateful to Lucy Campbell, Martin Davies, Robert Davies, Casey Doyle, Alexander Greenberg, attendees of the Oxford Philosophy of Mind Work in Progress seminar, and this journal's anonymous reviewers for their comments on earlier versions of this paper.

Since completing his DPhil at University of Oxford, **Ben Sorgiovanni** has been lecturer in philosophy at The Queen's College, Oxford. His research interests include epistemology, philosophy of mind, and Wittgenstein.

## References

- Audi, R. 1994. "Dispositional Beliefs and Dispositions to Believe." *Noûs* 28 (4): 419–35.
- Barnett, D. J. 2015. "Is Memory Merely Testimony from One's Former Self?" *The Philosophical Review* 124 (3): 353–92.
- Bernecker, S. 2009. *Memory: A Philosophical Study*. Oxford: Oxford University Press.
- Boyle, M. 2009. "Two Kinds of Self-Knowledge." *Philosophy and Phenomenological Research* 78 (1): 134–64.
- Boyle, M. 2011. "Transparent Self-Knowledge." *Proceedings of the Aristotelian Society* 85 (supp. vol.): 223–41.
- Boyle, M. 2017. "Self-Knowledge, Self-Concern and the First-Person Perspective" (manuscript).
- Boyle, M. 2019. "Transparency and Reflection." *Canadian Journal of Philosophy* 49 (7): 1012–39.
- Byrne, A. 2011. "Transparency, Belief, Intention." *Proceedings of the Aristotelian Society* 85 (supp. vol.): 201–21.
- Cassam, Q. 2014. *Self-Knowledge for Humans*. Oxford: Oxford University Press.
- Dancy, J. 2000. *Practical Reality*. Oxford: Oxford University Press.
- Debus, D. 2010. "Accounting for Epistemic Relevance: A New Problem for the Causal Theory of Memory." *American Philosophical Quarterly* 47 (1): 17–29.
- Dummett, M. 1993. *The Seas of Language*. Oxford: Oxford University Press, 411–28.
- Fernández, J. 2006. "The Intentionality of Memory." *Australasian Journal of Philosophy* 84 (1): 39–57.
- Frise, M. 2017. "Internalism and the Problem of Stored Beliefs" *Erkenntnis* 82 (2): 285–304.
- Hazlett, A. 2010. "The Myth of Factive Verbs." *Philosophy and Phenomenological Research* 80 (3): 497–522.
- Martin, C. B., and Deutscher, M. 1966. "Remembering." *The Philosophical Review* 75 (2): 161–96.
- Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.
- Moran, R. 2004. "Replies to Heal, Reginster, Wilson, and Lear." *Philosophy and Phenomenological Research* 69 (2): 455–72.
- Moran, R. 2012. "Self-Knowledge, 'Transparency,' and the Forms of Activity." In *Introspection and Consciousness*, edited by D. Smithies and D. Stoljar, 211–36. New York: Oxford University Press.
- Owens, D. 2000. *Reason without Freedom: The Problem of Epistemic Normativity*. London: Routledge.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Recanati, F. 2007. *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
- Rowlands, M. 2009. "The Concepts of Memory." In *The Routledge Companion to Philosophy of Psychology*, edited by S. Robins, J. Symons, and P. Calvo, 336–45. London: Routledge.
- Scanlon, T. 1999. *What We Owe to Each Other*. Cambridge, MA: Belknap.
- Schechtman, M. 1990. "Personhood and Personal Identity." *Journal of Philosophy* 87 (2): 71–92.
- Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Shah, N., and D. Velleman 2005. "Doxastic Deliberation." *The Philosophical Review* 114 (4): 497–534.
- Squire, L. 2009. "Memory and Brain Systems." *Journal of Neuroscience* 29 (41): 12711–16.
- Squire, L., and S. Zola-Morgan. 1988. "Memory: Brain Systems and Behaviour." *Trends in Neurosciences* 11 (4): 170–75.
- Tulving, E. 1972. "Episodic and Semantic Memory." In *Organisation of Memory*, edited by E. Tulving and W. Donaldson, 382–404. New York: Academic Press.
- Turri, J. 2009. "The Ontology of Epistemic Reasons." *Noûs* 43 (3): 490–512.
- Werning, M., and S. Cheng. 2017. "Taxonomy and Unity in Memory." In *The Routledge Handbook of Philosophy of Memory*, edited by S. Bernecker and K. Michaelian, 7–20. London: Routledge.

---

**Cite this article:** Sorgiovanni, B. 2020. The Role of Memory in Agential Self-Knowledge. *Canadian Journal of Philosophy* 50: 413–425, doi:10.1017/can.2019.51