Epistemic Value in the Subpersonal Vale

Abstract: A vexing problem in contemporary epistemology – one with origins in Plato's *Meno* – concerns the value of knowledge, and in particular, whether and how the value of knowledge exceeds the value of mere (unknown) true opinion. The recent literature is deeply divided on the matter of how best to address the problem. One point, however, remains unquestioned: that if a solution is to be found, it will be at the personal level, the level at which states of whole persons, as such, appear. We take exception to this orthodoxy, or at least to its unquestioned status. We argue that subpersonal states play a significant – arguably, primary – role in much epistemically relevant cognition and thus constitute a domain in which we might reasonably expect to locate the "missing source" of epistemic value, beyond the value attached to mere true belief.

Keywords: Subpersonal cognition; epistemic value; virtue epistemology; philosophy of cognitive science

I. Epistemic Value and the Swamping Problem

The past decade or so has witnessed a 'value turn' in mainstream epistemology (Riggs 2008). Of particular interest have been matters connected to *epistemic value*, value arising from the existence of distinctively intellectual goods, in contrast to, for example, moral or aesthetic goods (Kvanvig 2003; Pritchard 2007, 2011; Haddock, Millar, and Pritchard 2009; Baehr 2009). The investigation of epistemic value has generated various puzzles, one of which owes its existence partly to a compelling piece of received wisdom: that the epistemic value of knowledge outstrips that of mere true belief;[1] it may, for instance, be epistemically valuable to believe some true proposition <*p*> via a fortunate guess, but it is even more epistemically valuable to *know* that <*p*>. A subject who, for example, works stepwise through a proof of logical theorem T_{1L} and, as a result, correctly affirms T_{1L}'s status is in a better epistemic state than the person who unthinkingly spits out 'yes, it's a theorem' as a mere guess or only because, say, that person likes the font in which T_{1L} is set.

---

[1]In Plato's *Meno*, Socrates asserts (uncharacteristically) that he knows this, if he knows anything at all. See Kvanvig (2003, Ch. 1) for discussion.

And this is the case, to reiterate, even though *both* subjects end up with a true belief about T_{1L}'s status as a theorem.

This apparent truism makes mischief in the following way. Consider what would seem to be an equally plausible principle: whatever we say about the value of knowledge must be consistent with what we say about the *nature* of knowledge (Kvanvig 2003). If we accept that the epistemic value of knowledge exceeds that of mere true belief, we should not endorse an analysis of knowledge foreclosing that very possibility; rather, our analysis of knowledge must comport with an account of its constituents (e.g., justification or warrant) such that, plausibly, being justified (or warranted) adds value to what would otherwise be a merely true belief. But, some influential accounts of knowledge have struggled to satisfy this conditional constraint.

Take, for example, a straightforward process-reliability account of knowledge, according to which knowledge is type-identical to reliably produced true belief (Goldman 1979; Olsson 2007). If the process reliabilist maintains that the epistemic value of knowledge exceeds that of mere true belief, it is incumbent upon her to demonstrate just *how* the epistemic value of reliably produced true belief could exceed the value of mere (that is, not reliably formed) true belief. Here the reliabilist runs into trouble. Consider Zagzebski's telling comparison (Zagzebski 2003): an already good-tasting cup of coffee takes on no additional gustatory value simply in virtue of its being the product of a machine that reliably produces good-tasting cups of coffee; likewise, it is unclear how a true belief would become *additionally* epistemically valuable if turns out that the belief is not only true, but also the product of a reliable belief-forming process.[2] It seems that the epistemic value of truth "swamps" the value of reliability (Kvanvig 2003, 47–8). But – and here's the bad news for reliabilism – if the epistemic value of truth swamps the epistemic value of reliability, then it is false that the epistemic value of reliably formed true belief exceeds the epistemic value of mere (unreliably formed) true belief. Thus, the reliabilist's account of the nature of knowledge appears not to be consistent with the assumption that the epistemic value of knowledge exceeds that of mere true opinion.

And though this swamping problem (Kvanvig 2003, 2010, Zagzebski 2003, Jones 1997, Pritchard 2009b, Swinburne 1999, and Sylvan 2017) is often treated as an objection specifically to (at least standard forms of) reliabilism, it threatens other analyses of knowledge in like fashion (Kvanvig 2003, 2010; Pritchard 2009): in the case of *any* account of knowledge which aims to uphold the plausible assumption that the epistemic value of knowledge exceeds that of mere true belief, whatever conditions must, according to that account, be satisfied by knowledge must themselves be such that their satisfaction adds epistemic value to a true belief. And, virtually all otherwise promising theories of knowledge (or warrant, or justification) seem to flounder in the face of this challenge.[3]

---

[2]Kvanvig (2003) makes a parallel point in terms of the value of chocolate.

[3]Pritchard (2010) has suggested that, among existing contenders, robust virtue epistemology (e.g. Sosa 2007; Greco 2010) boasts resources most likely to vindicate the value of knowledge in the face of the swamping problem. However, at least as Pritchard sees it, robust forms of virtue epistemology are materially inadequate. And, perhaps more germane to present purposes, virtue-theoretic proposals are generally uninformative

Responses to the swamping problem typically take one of three forms: *validationist, fatalist*, or *revisionist* (Pritchard 2009a, 19–20). Validationists (e.g., Greco 2010; Goldman & Olsson 2009; Olsson 2010; Olsson 2007) argue that, when knowledge is understood properly, the epistemic value of knowledge is not swamped by the value of true belief. By contrast, *fatalists* (e.g., Baehr 2009; Ridge 2013) argue that we circumvent the swamping problem by rejecting the underlying intuition about the value of knowledge. Maybe, as this line of thought goes, knowledge is not as valuable as we initially thought it was. Finally, *revisionists* (e.g. Kvanvig 2003; 2010; Pritchard 2009; 2010; Riggs 2009) agree with fatalists that knowledge lacks a distinctive epistemic value, one not shared by mere true belief, while claiming that it is on closer inspection something *else* – typically understanding – which possesses the distinctive sort of epistemic value mistakenly attributed to knowledge.

Despite deep disagreement about how best to solve the swamping problem, all parties seem to agree where to look in conceptual space: if there is a solution, it is to be found at the personal level. What is the personal level? We say more about this matter in due course, but, in broad strokes, it is the realm of states of persons "as such, as experiencing, thinking subjects and agents" (Davies 2000*a*, 88, summarizing Dennett's original thought when introducing the idea of a distinctively personal level of explanation or description).[4] The standard formulation of the swamping problem presupposes a contrast between knowing and merely truly believing, and both belief and knowledge are widely thought to be states of whole persons as such. Moreover, these personal-level states are thought by many to have a distinctive sort of content; in fact, it is sometimes claimed that the kind of content had distinctively by personal-level states is the only genuine form of content and that only this sort of content has such genuine epistemic properties as *carrying justificatory force* (McDowell 1994*a*, 1994*b*). Generally, the consensus holds that personal-level states and relations between them provide the subject matter of epistemology and thus that, if a solution is to be found to the swamping problem, it will be found there – in the nature of personal-level states and relations between them or their contents.

We take exception to this orthodoxy, or at least to its unquestioned status. One of the present authors would go further, claiming that there is no personal level of reality *at all* (cf. Drayson 2012) and would also abandon the idea of a relatively autonomous and informative domain of personal-level explanation. On this level-eliminating view, standard folk-psychological states, including conscious states, may well exist, but if they do, they appear at the same level as, say, states of a face-recognition module or a language-parsing system; mind and cognition are, so to speak, flat – or flattened "from above," anyhow.

A less tendentious view, however, suffices to motivate and frame the remaining discussion: that subpersonal states and their content[5] play a significant role in epistemically relevant

when pitched at the personal level, a concern we develop in more detail below.

[4]Dennett introduced the distinction (1969, 93) as a difference in styles of explanation (or between sets of explanatory resources or vocabularies). In our discussion, we focus on the corresponding ontological questions – of supposedly personal-level states, properties, or processes – as is frequently done in the contemporary literature.

[5]Although we emphasize the role of so-called subpersonal-level states and take the content of such states

cognition – *particularly in cases in which there is no substantive model of what one might think should be the corresponding personal-level process* – and thus constitute a domain in which we might reasonably expect to locate the (or, at least *a*) missing source of epistemic value, beyond the value attached to mere true belief.

Two preliminary comments about our argumentative strategy are in order. First, although we elaborate below on the notion of a personal level, we do not take ourselves to be providing a definitive account of such a level; extant discussions of the personal level include many strands, and it is possible that they do not, taken collectively, express a single, unified conception of the personal level. For our purposes, it matters not, so long as the various authors who invoke the personal level would agree that the processes we appeal to, in order to solve the swamping problem, are, in fact, subpersonal. In other words, our arguments do not presuppose a unified or precise characterization of the personal level; a disjunction of possibilities will do. What matters is that the states and processes with which we are concerned count as subpersonal – not covered by any of the disjuncts in question – and here our claims seem incontestable. Thus, we can proceed to argue for the primary thesis of this paper without resolving disputes among proponents of personal-level theorizing regarding the proper characterization of the personal level. The kinds of processes we are interested in are subpersonal by any reasonable standard, and that suffices to support our claim that the swamping problem can be solved, in ways thus far overlooked entirely, by appeal to subpersonal processes.

Second, we do not press the following, relatively trivial point: that all personal-level states or processes are grounded in, enabled by, or realized by physical states or processes or supervene on physical states or processes (of the brain, it is typically thought, or the brain together with the immediate physical environment). A broad consensus in philosophy of mind and epistemology accepts some form of physicalism, and we do not intend merely to endorse that consensus. We have a narrower point in mind. We claim that, in some cases, personal-level states have epistemic value the source of which seems utterly mysterious viewed from the standpoint of the personal level. In such cases, there is no identifiable personal-level process or set of relations that might account for the value of such states. Yet, when one views the situation from the subpersonal perspective, the source of epistemic value comes into view. The matter might be best seen as involving a structural mismatch: only bare structure appears at the personal level, while a more richly structured process appears at the subpersonal level; as a result, the attempt at personal-level analysis falls short for want of explanatory resources, while, in contrast, subpersonal-level structure and processes offer plentiful explanatory resources, many of them relevant to questions of justification, or so we will argue. Note further that our arguments from structural mismatch

_____

to be particularly relevant to questions about epistemic value, we do not commit ourselves to a distinctive form of content at the subpersonal level. It might be that the content of the relevant subpersonal states is of the same sort as the content of supposed personal-level states, but that the state-types are different (*cf.* Heck 2000, which makes a parallel point regarding the conceptual-nonconceptual distinction). That is to say, we might solve the swamping problem by moving the discussion to the subpersonal level, though the content found at that level might or might not differ in its nature from the kind of content typically supposed to appear at the personal level.

take a twist. Personal-level processes lack not only positive, value-conferring structure; they also lack a relevant negative dimension. We identify subpersonal structure (with no personal-level correlate) that *prevents* some merely true beliefs from having as much value as justified true beliefs, by, for instance, preventing their continued existence.

Section two further motivates our focus on subpersonal states and processes, in contrast to those at the personal level. Section three relates such a shift in perspective specifically to extant proposed personal-level solutions to the swamping problem, arguing that such proposals show more promise when recast at least partly in subpersonal terms and supplemented accordingly. In section four, we offer two arguments directly in support of a subpersonal solution to the swamping problem.

II. The Personal and the Subpersonal

What might distinctively personal-level states or properties be? Perhaps no single author who writes about the personal level invokes every one of the properties listed below. Nevertheless, many of the following possibilities will likely sound familiar to the reader:

1. personal-level states are conscious;

2. personal-level states are states of whole persons, rather than of persons' bodies or body parts;

3. personal-level states are the states countenanced by folk psychology, such as beliefs, desires, hopes, and fears;

4. personal-level states appear in a space of reasons, as opposed to the space of mere causes;

5. personal-level states have an essentially normative dimension not captured by scientific descriptions of patterns of causally related events;

6. *a priori* reflection (on such concepts as *belief* and *desire*) or rational intuition can reveal the nature of personal-level states;

7. introspective acts can reveal the presence, nature, or causal contributions of personal-level states;

8. folk-psychological common sense can reveal the presence, nature, or causal contributions of personal-level states;

9. personal-level states, capacities, or abilities are the *explananda* of cognitive science, the latter's job being to account for the factors or mechanisms that implement or enable the appearance of the former;

10. personal-level states are such that attributions or descriptions of them are intensional (not extensional);

11. personal-level states are states of rationally coherent agents;

12. personal-level states rationalize action, as components of the belief-desire pairs that cause action;

13. personal-level states have genuine content, suited to the playing of a justificatory role, not possessed by states normally placed at the subpersonal level;

14. personal-level states have a kind of content that is transparent to, or even constitutive of, conscious experience – so-called phenomenal content.[6]

The preceding list is wide-ranging, and although clusters of entries bear deep relations to each, we make no attempt to convince the reader that this list represents a single, unified conception. Our fundamental task can be achieved so long as the status of the processes in which we're interested, *vis-à-vis* the preceding list, is "none of the above," that is, so long as such processes are paradigmatically subpersonal.

The personal level has taken center stage in epistemology. On some views, knowing is itself a primitive, personal-level mental state-type (Williamson 2000). According to a more firmly established tradition, a subject's knowing that *P* entails the subject's believing that *P*, because the belief-state is a component of the knowledge-state (but see Myers-Schulz and Schwitzgebel 2013 and Farkas 2016). And, of course, beliefs are typically treated as folk-psychological states the central facts about which are, as the tradition would have it, known by common sense, introspection, or *a priori* conceptual analysis (Lewis 1972, 256). Moreover, the contents of such states – whether it is the content of beliefs or irreducible knowledge states – are taken to be propositional. Given the character of the state-types themselves and the nature of propositional content, such personal-level states have, it is claimed, modal or normative properties absent from the merely natural, causal order (McDowell 1994*b*). For example, their natures are partly constituted by the inferences they license, which themselves are determined by states' propositional contents. These inferences, and the modal profiles of the relations between the contents that ground them, are thought to be fundamental sources of justification. Memories and perceptual states have also played important roles in epistemology; these, too, seem to involve attitudes that the mind, self, or person takes toward propositions. This is controversial in the case of perceptual states, but to the extent that it is, it is thought to be a barrier to the justification of beliefs by one's perceptual states. The worry runs as follows: If justificatory relations are relations between propositional contents of personal-level states, and perceptual states don't have propositional contents, how could beliefs be justified by perceptual states?[7]

---

[6]Davies (2000*a*, 88–90, 2000*b*, 46) lists many of these, as does Shea (2013, 1064–1065) and Frankish (2009, 90–91); on the messiness of the personal-subpersonal distinction, see Drayson (2012, 2014).

[7]This idea—viz. that the justification of belief is parasitical on certain logical relations among propositions—is described by Ernest Sosa (1980, 8) as the "intellectualist model of justification," which he takes to be a presupposition of both foundationalist and coherentist approaches to the structure of justified belief. Sosa thinks we should resist such a model not least because it can't very well account for the justificatory role of non-propositional experiences.

Epistemological internalism—and, in particular, accessibilist forms of internalism[8]—has seemed to make a particularly strong connection to the personal level. The role of the conscious appreciation of the content of mental states takes pride of place in the Cartesian epistemological project, as presented in the *Meditations*; on a standard interpretation of Descartes's strategy, the justificatory foundation of all of knowledge rests on (infallible) conscious access to the contents of one's own personal-level mental states. And, contemporary rationalists (e.g., BonJour 1998) continue to show deep sympathy for views in this vicinity. Consider, too, the phenomenal conservative's view (Huemer 2007): its seeming to the subject that *P* provides *prima facie* justification that *P*. Seemings are conscious (or intellectual) states that involve the subject's grasping of a content, which again places the source of justification at the personal level.

Nevertheless, much of what cognitive science has delivered over the past fifty years suggests a different picture of human epistemic activity, one that pushes our theorizing about human behavior into the subpersonal arena. It would be fatuous to anoint a single vision of human cognition as *the* result of cognitive-scientific enquiry. It is fair to say, however, that the further cognitive science has progressed, the greater the extent to which its results have marginalized the distinctively personal-level properties listed above. A wealth of results has, for instance, challenged prevailing assumptions about the human capacity for accurate introspection and the human ability to report accurately on the cognitive activity that produces one's own behavior (Nisbett and Wilson 1977; Wegner 2002; Wilson 2002; Lau et al. 2007; Haybron 2007; Schwitzgebel 2008; Alfano 2013; Harman 2000; Doris 2002; Gendler 2008).[9] Such work tends to undermine the claim that the revealing spotlight of consciousness plays a privileged role in our mental or cognitive lives, a role indicative of a distinctive level that should be of special interest to cognitive science. Accordingly, it would be a mistake to think that cognitive science's goal is to explain the mechanistic basis of the supposedly causally efficacious states or capacities revealed directly to human consciousness, as opposed to modeling directly the processes that produce intelligent behavior. After all, why should the job of cognitive science be to vindicate false or dubious claims about supposed personal-level causes of behavior? (Of course, modeling the cognitive processes that produce such false claims – where the issuing of the claims is itself taken as data – certainly *is* part of the job of cognitive science; Dennett 1991, Piccinini 2010). Given the state of the empirical evidence, then, it makes good sense for epistemologists to look to the subpersonal level. At the very least, epistemologists should recognize the possibility of joint contribution by so-called explicit and implicit processes to the production of behavior in the very cases in which belief and knowledge are thought to be relevant. If this messy collection of personal and subpersonal processes

---

[8]See Chisholm (1977) and Bonjour (1985, ch. 2) for classic defences of the accessibilist position; for a recent overview, see Madison (2010).

[9]A reasonable response to much (though not all) of this work is to claim that subjects have highly accurate introspective access to their most immediate experiences, that is, to what they are thinking or experiencing *right now* (Gertler 2011). This might be correct but still represents only a Pyrrhic victory for the personal-level introspectionist; the accuracy of such introspective assessments appears to be too short-lived to be of much epistemic use.

jointly produce the behavior to which matters epistemic are deemed relevant (Dennett 1991; Perugini 2005; Rey 2001; Gendler 2008; Gawronski and Bodenhausen 2014) – and not in such a way that the subpersonal-level contributors could plausibly be characterized as mere realizers of the personal-level contributors – then one might reasonably suspect that epistemically relevant properties attach both to subpersonal and personal processing streams.

Here is where we have gotten to. Philosophers have tended to think of epistemology as a personal-level enterprise and only a personal-level enterprise; it's about beliefs, memories, testimony, perceptions, and their contents, and justification-relevant relations between them. Moreover, matters epistemic, and personal-level matters more generally, are thought to bear significantly on our behavior – on what we do.[10] But, our best attempts to understand the wellsprings of behavior call into question the supposedly privileged role of personal-level states with respect to the production of behavior; instead, cognitive science seems to reveal a range of subpersonal processes that produce much of the behavior historically thought to be produced by personal-level states, including by justified beliefs. To a great extent, then, subpersonal processes account for the instrumental value associated with our personal-level cognitive successes. When our beliefs produce behavioral success, it is by working in tandem with subpersonal processes *that are not implementations of those beliefs*. Perhaps subpersonal processes also distinctively ground the *non*instrumental value of personal-level states.

To be clear, we take it to be of interest to show, in the case of merely true beliefs, how their instrumental value might be enhanced by the accrual of justification (or warrant) via distinctively subpersonal-level processes (that is, processes not mirrored by processes at the personal level); and some of what we say in the sections to follow may contribute to this project. It is more challenging to solve the swamping problem with regard to noninstrumental value, but we will try to show that a subpersonally oriented strategy holds promise in this regard as well, particularly in cases of a mismatch between rich, epistemically relevant subpersonal processing and relatively impoverished structure at the personal level.

III. Subpersonal Transformations of Personal-Level Proposals

In this section, we argue that an emphasis on subpersonal-level processing improves the prospects of two proposed personal-level solutions to problems concerning the value of knowledge. Section III.A argues that one extant reliabilist attempt to solve the swamping

---

[10]A testament to this point is the contemporary literature on the normativity of action and practical reasoning (e.g., Fantl and McGrath 2002; Hawthorne 2004; Stanley 2005; and Hawthorne and Stanley 2008; Gerken 2011; 2014; Williamson 2000; *cf.*, Simion *forthcoming*), which features in the main discussion of *which* personal-level states (e.g., knowledge, justified belief) best rationalize action.

problem – viz. Olsson's (2007) argument from increased practical value – must appeal primarily to subpersonal processes the image of which is not manifest at the personal level, if the proposal is to be viable. In Section III.B, we consider what is taken to be the most promising personal-level response to the swamping problem currently on offer – *viz.* the solution offered by (robust forms of) virtue epistemology (e.g., Greco 2010) – and argue that a subpersonal variation on this approach has more to recommend it.

III.A

Erik J. Olsson (2007) argues that "reliabilist knowledge promotes successful action over time...[because]...reliabilist knowledge promotes stability and...stability is conducive to successful action over time" *(ibid.,* 349). According to Olsson, when an unreliable process produces a (mere) true belief that *P*, the very unreliability of that process will likely undercut or neutralize, eventually, that belief's potential to contribute to successful action. A given subject deploys a given mechanism or runs a given process-type (we treat these as equivalent for present purposes) repeatedly over the course of her life. In the cases in question, the process-type is, by hypothesis, unreliable; thus, the preponderance of later applications of it – that is, those that occur after the time at which the application of the process led to the fixation of the subject's true belief that *P* – will yield false beliefs. As Olsson sees things, subjects track the sources of their belief, recording which processes produce which beliefs as well as the rate of past success and failure of various processes to produce beliefs that lead to effective action. Thus, a subject who continues to use the process in question will subsequently doubt or reject *P*; given feedback from the world, the subject will detect the falsity of the outputs of the majority of later applications of the process that produced *P*, which results will call into question *P* itself (even though *P* is, in fact, true), thereby robbing the true belief that *P* of what would have been its contributions to successful behavior – presumably because the subject abandons, or at least take a highly qualified attitude toward, the belief that *P*, and thus does not act on it.[11]

To be clear, to the extent that Olsson's tack succeeds, it does not provide what many would want from a response to the swamping problem: an account of why a justified true belief that qualifies as knowledge has more *final* (i.e., noninstrumental) value than a mere true belief. Rather, it focuses on a particular kind of instrumental value, *viz.* the value that

---

[11]This work builds on a simpler idea (Goldman and Olsson 2009) that having a reliably produced true belief is better than having a true belief produced by an unreliable process, because one's having a reliably produced true belief probabilifies one's having true beliefs in similar circumstances in the future. While this may be correct, it does not seem to increase the value of any individual belief. A true belief's having been produced by a reliable process entails the presence of a valuable tool in the subject, a tool such that, if the subject continues to possess and deploy it, it will produce a preponderance of true beliefs in the future. It is not clear, however, why the value of the possession of that tool would increase the value of a given belief produced by it, beyond the value the belief has in virtue of its truth. Instead, a belief's having been produced by a reliable process seems to be merely an indicator that the subject possesses a tool to produce true beliefs reliably (in certain kinds of circumstances); the relational fact of a belief's having been produced by a reliable process would seem to confer on that belief only a diagnostic role.

a justified true belief has in virtue of its being likely, itself, to continue to contribute to successful action. Nevertheless, the account does generally make sense of the intuition that justified true belief is more valuable than (mere) true belief, given the incredible importance of successful action in human lives.

Assume, for the sake of argument, that a proposal along the lines of Olsson's holds promise. Whence does it draw its explanatory power, the personal or subpersonal level? As noted above, Olsson supposes that agents generally record which beliefs were produced by which mechanisms and keep a running success rate of each mechanism; in Olsson's words, this "requires that the agent maintain a mental record, a record in her mind, of how beliefs were acquired" (Olsson 2007, 352). After all, if the agent were not successfully deploying such tracking abilities, she would not infer the likely falsity of P from later failures of the P-producing process to produce beliefs that support successful action. As Olsson conceives of things, then, the process in question plays out at the personal level.

A fleshed-out version of Olsson's story would seem to require that, relative to each use of a given belief-forming mechanism, the subject accurately encode, not just *that* the mechanism produced the belief in question, but also the context of such production, at correct level of specificity. Too often, a mechanism that has a weak track record in a context described at one level of specificity (use of vision while beneath the surface of the water in a naturally formed lake) has a stronger track record relative to a more inclusive set (use of vision *simpliciter*), or vice versa – points familiar from discussions of the Generality Problem for process reliabilism. Assume a given true belief P was produced by a mechanism in a context that would be appropriately individuated, for the purpose of determining P's level of justification, at a fine grain, and that relative to such a context, the mechanism is in fact unreliable. Let us say, too, that the subject mistakenly individuates, in her record-keeping process, the context in question in a coarser-grained way relative to which the mechanism in question is, in fact, reliable. In this case, the subject will *not* weed out the unjustified belief that P, because she will treat the contexts of the application of the mechanism in question in a coarser-grained fashion and will not come to see – on the basis of negative feedback from the world in the relevant finely individuated contexts – the mechanism that originally produced the belief that P as unreliable. If this is a relatively common phenomenon – if the subject doesn't identify and record contexts at the correct grain – the subject's various unjustified true beliefs may well persist and continue to contribute to successful behavior, *contra* Olsson's prediction; the subject will think of P as the result of the operation of a reliable mechanism (the visual system applied to the general spatial layout, for example), which generally gets things right, instead of seeing the mechanism as being applied in more finely individuated context (the visual system as applied in poor light to objects at a distance moving quickly), and thus won't abandon her unjustified belief that P.[12] Similarly, without accurate records of the sort in question, the subject may well treat a justified true belief as unjustified and abandon it in accordance with Olsson's

---

[12]This kind of concern is not merely theoretical. For instance, the sort of source information that some languages encode syntactically marks sources only at a very coarse grain – distinguishing between such categories as having been acquired by testimony or having been observed first-hand (Tosun et al. 2013).

schema. Thus, absent a commitment to the reasonably accurate tracking and recording of justification-relevant contexts of belief-formation, one should doubt that Olsson's schema identifies the distinctive source of value attached to justification.

How plausible is it, then, that human subjects track, at the personal level, the output of belief-forming mechanisms in contexts? Below we survey the relevant empirical literature, but let us be clear, from the outset, about the sort of evidence we should look for on Olsson's behalf. It should show that, at the personal level, (a) subjects track the sources (and contexts – take this as read in what follows) of their beliefs, (b) subjects do so reasonably accurately, (c) subjects use that information to calculate reasonably accurate track records of their various belief-producing mechanisms, and (d) subjects bring the results of those calculations to bear on commitments to past products of their belief-producing mechanisms (for example, to bear on the judgment that a past doxastic output of a given mechanism is likely to be true). This sets a very high bar, to be sure, and, perhaps unsurprisingly, the empirical literature runs in the opposite direction, supporting at least a moderately pessimistic view about every one of these desiderata and a firm skepticism concerning their joint satisfaction. The literature in question is enormous, but we shall, in what follows, fairly convey a sense of the obstacles faced by Olsson's personal-level account.

Consider first the so-called illusory-truth effect (Hasher et al. 1977; Dechêne at al. 2010). Mere exposure to sentences increases the likelihood that subjects will judge them to be true when asked later to make a judgment concerning their validity (that is, when asked whether the statements are true). This occurs even in cases in which subjects know better, that is, even when they have stable beliefs that contradict the information to which they're being exposed experimentally (Fazio et al. 2015). The leading explanation of this phenomenon appeals to a more general and well-established subpersonal construct, ease of processing, in this case created by previous exposure (Begg et al. 1992). Further results reinforce this hypothesis: merely setting (unfamiliar) sentences in an easier-to-process font increases the probability (over sentences set in a more difficult to read font) that subjects will judge them to be true after a single previous exposure (Reber and Schwartz 1999). In such cases, subjects simply are not attending accurately to the sources of their beliefs or to the mechanisms producing them. If they were aware, at the personal level, that a mere ease-of-processing mechanism were producing the beliefs in question, drawn only from, for instance, written text with no attached credibility, subjects would, presumably, not make the judgments they do.

The illusory truth results are by no means outliers. Empirical work documents various kinds of cases in which subjects fail to track the sources of their beliefs and mechanisms that produced them (see Marsh et al. 2008, for a review). In list-learning paradigms, for example, subjects exposed to a list of semantically interrelated words judge, when asked after exposure about words that would have "fit" onto the list, that the latter words were listed. They seem to have both a false belief and to make a false judgment about the source of that belief, thinking they heard or read the word when it was instead self-generated (by a process of semantic association). And, in some versions of the experiments, the personal-level record is strikingly corrupted; subjects report phenomenological experience

– a rich episodic memory – of, for instance, the experimenter having read the nonlist word aloud, even though it is only a lure and was not read aloud (Roediger and Gallo 2005; see Geraci and Franklin 2004 for cases in which subjects are misled by nonsemantic linguistic relations). In eyewitness suggestibility experiments, subjects report having witnessed what are actually false details that experimenters have in one way or another exposed the subjects to after the witnessing of the actual event in question; subjects confuse testimony-based belief-formation (verbal, in print) for first-person observations made at, for example, the scene of a car accident (Loftus 1979). A different line of research shows that déjà vu can be induced experimentally, by means of mere exposure, which again seems to be a mistake about sources (Brown and Marsh 2010). Subjects are susceptible to the false fame effect, incorrectly categorizing faces as being those of famous people, as result of mere past exposure to said faces in experimental settings (Jacoby, Kelley, Brown, and Jasechko 1989; Jacoby, Woloshyn, and Kelley 1989). And, subjects are more likely to choose the wrong subject from a line-up when they've seen that person's face in a book of "mug shots" prior to viewing the line-up (Brown et al. 1977). Subjects also make source errors when recalling the factors that influenced their decisions, in a way that systematically supports decisions made (Mather et al. 2000), attributing, for example, positive features that support the decision they made to the option chosen, when those positive features were actually attached to the option not taken. In addition, the effects of social contagion powerfully distort memory of sources (Meade and Roediger 2002; Barnier et al. 2008). And bear in mind that in every one of these cases, the measures used involve at least some (and often exclusively) personal-level judgments.

Marcia K. Johnson and collaborators (Johnson et al. 1993) developed the leading theory of the monitoring of sources of information and memories. Put simply, Johnson's source-monitoring framework grounds subjects' ability to identify the source of a given memory in the ability to associate the content of the memory itself with various cues and features of the context in which the memory was formed. Viewed in that light, the various results reviewed above may seem unsurprising, given the general remember-to-know (or, R-to-K) shift (Barber et al. 2008; Dewhurst et al. 2009) exhibited by human memory: the general tendency for information about the specific circumstances to be lost over time and for the supposed knowledge to be represented context-free, at least insofar as this can be probed at the personal level. This distinction parallels (and may be largely coextensive with) the widely made distinction between episodic and semantic memory (Tulving 1972) and the tendency of memories for general information about the world to shift from episodic form – replete with details concerning the context in which the information was acquired – to the semantic form, which encodes the information itself stripped of such details.

To be fair, some experimenters have gone to significant lengths to try to warn subjects about credibility or to get them to use information about credibility to mitigate false various memories (Echterhoff et al. 2005; Henkel and Mattson 2011; Chambers and Zaragoza 2001; Meade and Roediger 2002; Begg et al. 1992), which might seem particularly relevant to (c) and (d) above. These efforts are not wholly without results, but neither do they instill much confidence in human abilities or tendencies. To the extent that such warnings have

salutary effects (reducing eyewitness suggestibility effects, for example), the effects are weak, depend on the specific choice of wording (it should be given in the indicative, not the subjunctive), and on the timing of the warning (they are much more effective if they're provided before the "false testimony" not after). And, even in the specific cases in which the warnings work well, for instance, eliminating the effects of misinformation, this amounts only to the treatment of planted information in the same way as new information (including only at the time of later memory tests). But, given how badly subjects perform on newly introduced material (their high rates of false positives), this is not much consolation, partly because explicit memory for sources continues to be poor. Moreover, the salutary effects are relatively short-lived (Chambers and Zaragoza 2001, 1122). This doesn't jibe well with Olsson's framework; if his story is to be plausible, subjects must experience some general sense of failure of an action guided by belief produced much later by the mechanism, and they must recall that the mechanism in question produced a belief they had previously thought to be true. To the extent that only prewarnings have a sufficiently powerful effect, Olsson sacrifices that gain. For on his framework, the "warning" the subject gets is likely to come well after the fact.

The results on source credibility should be especially troubling to Olsson. In some cases, subjects' beliefs about the credibility of the source of an individual piece of information significantly affect their judgments: when they believe that a piece of information is from a credible source, they're significantly more likely to judge it true. But, even then, they have such poor memory for the actual sources of individual memories (Begg et al. 1992, 452) that, statistically speaking, the actual credibility of the source is not correlated with the subjects' pattern of endorsements (Henkel and Mattson 2011, 1708). Moreover, subjects continue categorize as true approximately 50% of statements they believe (rightly or wrongly) to be from a noncredible source (Begg et al. 1992, 451–453). The picture that emerges is of subjects who are not completely insensitive to considerations of credibility, but who have *(a)* mediocre source memory, *(b)* often make poor use of credibility information they have, and *(c)* who continue to make widespread errors concerning other matters to do with sources. Considering the multi-step, statistical nature of the procedure Olsson demands, such shortcomings compound. As a result, there's no reason to believe of any particular mechanism in a given subject, that the subject will have a sufficiently accurate record of the performance of that mechanism and will take it into consideration in the formation of new beliefs or the continued endorsement of belief previously produced by that mechanism. The breadth and the depth of subjects' mistakes – their failure to mark sources at all, their failure to mark sources accurately, their failure to use accurate information that they have – puts paid to Olsson's claims about personal-level record keeping in humans.

How promising, then, is a subpersonal reading of Olsson? We must concede that, given the determination of the personal by the subpersonal, every personal-level failure is, in some sense, a subpersonal one. It's worth noting, however, some reasons for tempered optimism about a subpersonal version of Olsson's approach. Cognitive scientists have documented various ways in which the cognitive system tracks temporal patterns and sources, for example, in the Iowa Gambling Task (as performed by normal subjects – see

Bechara et al. 2005) and in tasks that exhibit frequency effects (Jones et al. 2013). Also, the grammatical marking of sources of information can affect the accuracy of memory; in some languages, differences in, for example, verb inflections mark whether the event being reported was seen by the speaker or, instead, learned about by testimony, which effects, for example, one's ability to remember items marked as the result of first-hand experience (Tosun et al. 2013), and this suggests that subpersonally the subject tracks at least some of the "sources" of her beliefs.[13]

Note the extent to which our point here reflects our general argumentative strategy. We are not merely describing the mechanisms that implement a value-creating, personal-level process. Rather, our claim is that there is no such personal-level process. We thus find a structural mismatch between personal-level and subpersonal-level processes such that the source of value comes into view only at the subpersonal level.[14] Consider the Iowa Gambling Task, for instance, in which subjects receive rewards and penalties for drawing cards from a variety of decks; some decks have a much more profitable structure than others. For instance, in one deck, 10 out of 11 cards pay out $50 per card, while the eleventh shows a loss of $250, which equals an expected utility of approximately $23 dollars per draw. In another deck, 10 out of 11 cards pay $100 per card, while its eleventh shows a loss of $1250, which equals an expected utility of approximately -$23 per draw). Many subjects eventually achieve a conscious awareness of the superiority of winning decks over losing decks, but prior to that point (if it ever comes), subjects have no such awareness yet nevertheless systematically modulate their selections in favor of the winning decks (while also showing physiological signs of a sensitivity to threat of loss when beginning to reach for losing decks). Thus, a subpersonal tracking process takes place independently of any personal-level process whatever, for no parallel personal-level process occurs.

We acknowledge the limited scope of the evidence of subpersonal source tracking and evaluation. In contrast, the wealth of evidence against a personal-level story of the sort Olsson has in mind seems damning We close this subsection, then, with a conditional conclusion: If Olsson's strategy pays off at all, it will do so as a subpersonal-level account of processes that produce instrumental value; justified true beliefs are more valuable than merely true beliefs because the operation of certain forms of subpersonal processing increases the likelihood that a subject will continue to act on a true belief when it's produced

---

[13]In the experiments of Tosun (Tosun et al. 2013), subjects receive instructions meant to reduce personal-level attention to credibility-related questions: "To make the study phase more similar to a natural language situation in which participants would not be attempting to remember the sentences or the source of evidence, participants were told that the experiment was about their ability to comprehend sentences and their reading times would be measured." (*ibid.*, 125)).

[14]To reiterate, our point is *not* that two processes run in parallel, a personal-level process and a subpersonal level one, and that the subpersonal process is the more metaphysically or explanatorily fundamental of the two, providing the genuine explanation of epistemic value, in opposition to a parallel personal-level explanation that might also appear to be able to do the job. Rather, our point is that *no* personal-level process of the appropriate sort exists; such a process appears *only* at the subpersonal level. Thus, our criticism of Olsson's personal-level proposal in no way rests on any claim about causal-explanatory exclusion or the relation between realizers and realized states or between supervening properties and their supervenience base.

by a reliable (and thus justified) process, as compared to beliefs produced by unreliable processes.

III.B

In III.A, we argued that Olsson's proposal has promise only insofar as it appeals primarily to claims about the subpersonal level; for, that seems to be where the value-creating action appears, if it appears at all, in the manner in which Olsson claims it does. In apparent contrast, virtue epistemology seems to have the resources to articulate a cogent, personal-level solution to the swamping problem.[15] According to a virtue-based approach, knowledge is true belief the correctness of which is *because of*, or which manifests intellectual virtue on the part of, the agent (Zagzebski 1996; Sosa 2009; Greco 2010; Haddock, Millar, and Pritchard 2010). In this section, we argue, however, that a subpersonal variation on the virtue epistemologist's proposed solution to the swamping problem fares considerably better than an exclusively personal-level account.

For ease of exposition, we focus on John Greco's (2010) canonical presentation of the virtue-theoretic response to the swamping problem. According to Greco, knowledge is a cognitive success (i.e., the attaining of a true belief) that is because of cognitive ability. Furthermore, *achievements* are defined more generally as successes that are because of ability. Thus, knowledge is a cognitive achievement, the achievement of a true belief reached through ability. Achievements are valuable for their own sake (in a way mere lucky successes are not); therefore, *knowledge*, qua achievement, is valuable for its own sake.[16]

Let us grant for the sake of argument that knowledge is a cognitive success because of cognitive ability, and thus, that knowledge is always and everywhere a kind of achievement (it is success because of ability). Even on these assumptions, the thesis that knowledge is valuable for its own sake, in a way that mere (unknown) true belief is not, follows only if being reached *through ability or virtue* suffices to make a true belief more valuable than its nonknown counterpart. But why should this be?

At this juncture, Greco takes a nod from Aristotle. In the *Nicomachean Ethics*, Aristotle distinguishes between achieving an end through luck and achieving the end through the exercise of one's abilities (or virtues). The latter, according to Aristotle (as Greco 2010 summarizes):

> is both intrinsically valuable and constitutive of human flourishing … In this discussion Aristotle is clearly concerned with intellectual virtue as well as moral virtue: his position is that the successful exercise of one's intellectual virtues is both intrinsically good and constitutive of human flourishing (2010, 97-98).

---

[15]This point has been conceded in various places by Duncan Pritchard (e.g., 2010; 2009a; 2009b), who is a leading critic of robust virtue epistemology.

[16]This is a condensed version of the argument found in Greco (2010).

We think that a variation on Greco's response that shifts the discussion to the subpersonal level can do better than an exclusively personal-level account. The claim that "the successful exercise of one's intellectual virtues is intrinsically good" is put forward as an explanation for why knowledge, conceived of as a kind of successful exercise of intellectual virtue, is valuable in a way that unknown true belief is not. But to say that successful exercise of intellectual virtue is *intrinsically* valuable means just this: that exercising one's intellectual virtues is good for its own sake in virtue of properties that are internal to the successful exercising of intellectual virtue.[17]

But what *are* these properties? Identifying these properties involves locating a *supervenience base* that is intrinsic to the successful exercising of intellectual virtue.[18] Without actually locating such a supervenience base in the successful exercise of intellectual virtue, the virtue-theoretic answer leaves a sizeable explanatory remainder.[19]

Indeed, even if there is some pull to thinking that intellectual achievements are valuable for their own sake, the strategy of appealing to the very *fact* of something's being intrinsically valuable (in an explanation for why the thing in question has the value it does) begins to look like a case of *obscurum per obscurius*. Even if something is valuable for its own sake – and thus, finally valuable – because of its intrinsic properties rather than extrinsic properties, we want to know *what those properties are*, not *just* that they are properties internal to the successful exercise of virtue. What is the *supervenience base* of this intrinsically valuable thing? In virtue of exactly what properties internal to the successful exercise of virtue is the successful exercise of virtue valuable for its own sake? On the account offered, the answer seems to be: because Aristotle says so.[20]

To be clear, our point is not simply that the successful exercise of virtue has an interesting

---

[17]This way of thinking about intrinsic value owes originally to Moore (1903). For a more recent extended discussion, see Rabinowicz and Ronnow-Rasmussen (2000).

[18]Note that the supervenience base of an object's value should not be confused with the related idea of the *constitutive grounds*. See Rønnow-Rasmussen (2011).

[19]The answer is tantamount to the following: "Knowledge is valuable (in a way that unknown true beliefs are not) because it involves something that is (*à la* Aristotle) intrinsically valuable and which is not shared by mere true beliefs: the successful exercise of intellectual virtue. Put another way, knowledge involves something that is valuable for its own sake in virtue of properties intrinsic rather than extrinsic to the exercise of virtue. But what these properties actually are we shall not say—we say only that they are internal to the successful exercise of intellectual virtue."

[20]We do not here rely on the excessively strong claim that any explanation failing to mention a supervenience base leaves the kind of sizeable explanatory remainder in question. Rather, the problem at hand results from a wide gap between the virtue-theorist's personal-level explanation, which we argue is thin – "you succeeded because you have the sort of virtuous character that allows you to succeed!" – and the explanation one finds when one examines the supervenience base, which is rich. In contrast, in some cases, we may find a rich personal-level explanation the essential structure of which is reflected in the structure of its supervenience base. In such a case, knowledge of the supervenience base can help to confirm the accuracy of the personal-level explanation; and so long as we have good reason to believe in a personal level, facts about the supervenience base do not supplant the personal-level facts as the source epistemic value; for, in that case, the relations present in the supervenience base do not distinctively account for the value in question.

supervenience base or that it is grounded in a set of interesting subpersonal, implementing processes. That would be hardly to go beyond a kind of physicalism presupposed by nearly all contemporary analytic epistemologists. Rather, we propose that a move to the subpersonal level provides a better answer – a richer answer that illuminates the source of the value in question, in contrast to the relatively empty personal-level virtue-theoretic proposal on its own. This point can be thought of as another instance of "mismatch" reasoning. The virtue-theoretic, personal-level story offers little structure to work with or little in the way of a fleshed-out account of the value-determining properties that might be of use to those trying to understand how the distinctive value of knowledge is constructed or in what it consists. The contrasting, subpersonal story is more illuminating because the relevant subpersonal-level states and processes comprise additional pertinent structure.

Consider one way a virtue epistemologist might add additional meat to the personal-level account of epistemic value. Intellectual virtues, as such, must be truth-oriented dispositions that are appropriately *cognitively integrated* (e.g., Pritchard 2010; Greco 2010, 156, *passim*) within the agent's cognitive character, a point that is embraced elsewhere by Greco himself when distinguishing true beliefs reached through virtues from true beliefs reached through reliable but 'strange and fleeting processes', the exercise of which issue beliefs that, even when true, fall short of knowledge.[21] Cognitive successes that involve "the successful exercise of intellectual virtue" (i.e., that which the virtue epistemologist tells us is intrinsically valuable) are thus cognitive successes the formation of which is grounded in truth-oriented dispositions that are stable and *integrated*, as opposed to being merely fleeting or disintegrated. The properties of a truth-oriented disposition in virtue of which it is cognitively integrated within the agent's wider cognitive character are thus properties in virtue of which it is valuable. What are *these* properties?

These appear to be subpersonal properties. Here we consider two ways in which cognitive integration, of the sort adverted to by the virtue epistemologist, appears as a subpersonal phenomenon. The first involves the very nature of cognition and its relation to integrated cognitive systems. The second involves processes by which individual states or abilities become integrated into an existing cognitive system.

Questions about cognitive integration and cognitive systems have arisen forcefully in the recent debates in the philosophy of mind, particularly in connection with the Extended Mind Hypothesis (EMH) (Clark and Chalmers 1998) and the proposal that groups sometimes constitute cognitive systems (Hutchins 1995, Huebner 2013). A prominent thread in the debate over EMH can be summarized as follows: To the extent that Clark and Chalmers consider personal-level states (in the context of, e.g., their discussion of Otto and his notebook), their attempts to support the extended view bog down (Rupert 2004, 2009, 2013). If, as they indicate, they wish to support the extended view of the mind by appeal to its causal-explanatory superiority – one that privileges natural kinds typed coarsely

---

[21]See here Greco's (2010, 156) diagnosis of Plantinga's (1993) brain lesion case. For a related discussion of cognitive integration and its connection with virtue epistemology, see Pritchard (2010) and Menary (2012). For discussion of cognitive integration in the context of the extended mind debate, see Menary (2010).

enough to include a significant number of real-world instances with partly external minimal supervenience bases – there must be a successful science of personal-level cognition that individuates cognitive state-types very coarsely. But, there's very little extant science of this sort; cognitive science tends to produce fairly fine-grained models.

At this juncture, one naturally turns to cognitive science in search of a boundary that distinguishes cognitive from noncognitive causal contributors to the production of intelligent behavior. One such strategy appeals to the line between causal contributors that appear within the relatively persisting, relatively integrated cognitive system and those that appear beyond it (Wilson 2002, Rupert 2009). This requires, however, some specification of a measure of integration. Rupert appeals to various conditional probabilities of mechanisms' co-contribution to the production of the subject's intelligent behavior. The view is not without its problems (Klein 2010, de Brigard 2017), and competing proposals have been made. For example, drawing on the work of Sporns and his colleagues (Sporns, Chialvo, Kaiser, and Hilgetag 2004), Goldstone and Gureckis appeal to a measure of computational complexity (Goldstone and Gureckis, 2009, 428; also see Clark 2009, 251 n24) to characterize the sort of integration characteristic of cognitive systems. And, Edwin Hutchins proposes that a steep drop-off in the computational gradient ("steep gradients in the density of interaction among [representational] media") marks the boundary of the cognitively relevant unit of analysis (Hutchins 1995, 157). For present purposes, we emphasise only that, as it has taken shape, the debate clearly concerns subpersonal-level properties of the cognitive system. To the extent that progress is being made on the issue, it is only where contributors "descend" to the level of subpersonal processes and appeal to quantifiable relations between components of the cognitive system that are not consciously accessible (and are not personal level by any other influential standard).

Now consider a different sort of integration, the way in which newly acquired skills and memories are integrated into the subject's cognitive profile. One especially striking stage of the integration process occurs during sleep. Sleep consolidates skills and memories, and it does so in a way that allows a new motor routine or the content of new experiences to be incorporated into the subject's or cognitive system's overall functioning; this is partly a matter of maintaining balance with and facilitating behavior-controlling co-operation with other bodily skills and other parts of one's store of memories. The latter case often goes under the heading of memory consolidation, a central and relevant aspect of which is described by Dudai, Karni, and Born (2015): "Consolidation is a dynamic, generative, transformative, and lingering process that is posited to balance maintenance of useful experience-dependent internal representations of the world with the need to adapt these representations to the changing world" (21). And, this consolidation process is of particular importance in the case of propositional knowledge: "There is also growing evidence that this sleep-associated redistribution of information is accompanied with an increased semantization of memories and the abstraction of gist information from episodic representations" (*ibid.*, 23), and "the hypothetical process of systems consolidation is most commonly discussed within the context of declarative memory" (*ibid.*, 26). Moreover, the reconsolidation process appears to be directed at, and triggered by the need to, integrate

declarative memory into existing bodies of represented facts, of the sort relevant to inference: "Hence, one may hypothesize that, instead of external cues, reactivated pre-existing schemas in neocortical sites direct sleep-dependent consolidation, for example, by favoring the hippocampal reactivation of that memory information that fits the preexisting schema" (*ibid.*, 25), "[P]rior knowledge schemata shape the engagement of the hippocampus in declarative consolidation..." (*ibid.*, 26), and "At the same time, one should not overlook the postulated role of consolidation in balancing stability and change and maintaining adaptive predictive power of representations" (*ibid.* 28).

These may be strange-sounding processes, but their basis is not fleeting; it reflects fundamental operations of the subpersonal cognitive architecture. Furthermore, the processes in question clearly are not at the personal level, by any of the extant conceptions of the personal level listed above. And, notice that it's not simply a matter of cementing memories or practiced routines; it's simultaneously the maintenance of all of the subject's existing cognitive activities and skills; for the incorporation of anything new into that system involves the careful adjustment of relations among existing structures as well as relations to new ones, so as to maintain the integrated functioning of the entire system.[22]

In conclusion, if we wish to understand the kind of cognitive integration appeal to which fills the large explanatory lacuna in the virtue-epistemologist's proposed solution to the swamping problem, we do best to look—particularly at the point where Greco's explanation bottoms out—to the subpersonal level. Only there, it seems, do we find the structure and complexity that adds significant explanatory power to the virtue-epistemologists appeals to integration, such structure and complexity as sheds light both on what constitutes the appearance of a single integrated set of cognitive virtues and how such integration is dynamically maintained within a single cognitive system, in response to new information or pieces of evidence. On the virtue-based account, value flows from the exercise of intellectual virtues, which must be understood as part of an integrated psychology, integrated with regard to knowledge structures and to cognitive abilities. Such psychological integration is, however, a subpersonal matter.

IV. Straightforwardly Subpersonal Processing

In the current section, we argue directly for subpersonal solutions to the swamping problem, *sans* any detour through extant, personal-level accounts of the value of knowledge. Presently, we develop two arguments each of which rests on a subset of the following background assumptions:

---

[22]For more on memory consolidation, see Rasch and Born (2013) and Squire et al. (2015). Approaching the issue of integration from a slightly different angle, consider the problem of catastrophic interference, which afflicts many neural network models of learning and remembering; this problem arises when the changes in weights involved in the storage of a new pattern "overwrite" weights that encode previously stored patterns or associations. See Ans et al. (2004) and Srivastava et al. (2014) for (clearly subpersonal) attempts to solve this problem in what might be neurologically realistic ways.

Correspondence: In order for a belief to be true, its propositional content must correspond to reality.

Compositionality. It is a necessary condition on a belief's being a belief that *P* that the belief's components represent, express, or refer to the individuals, relations, properties, etc. constitutive of *P*.

Belief Endurance. In order that a subject's belief be the same belief at two points in time, it must be an attitude toward the same proposition at both of those points; more generally, for any two beliefs B1 and B2 (separated in whatever way) to be type-identical beliefs, B1 and B2 must be attitudes toward the same proposition.[23]

Subvenience. It is a nomologically necessary condition for a belief's continuing to be held that the subject who holds it continue to be in some subpersonal state(s) or other that subvenes (or realizes) the belief in question.

Ongoing State: Although the *acquisition* of a new belief might be an *event with a terminus*, having a belief is itself an *ongoing state* (Vendler 1957).

Maintenance: Having a justified belief is not only a matter of having a belief acquired under appropriate circumstances; it is also a matter of sustaining that belief in an appropriate way.

A few words are in order regarding these assumptions. Firstly, note that *Correspondence*, *Compositionality*, and *Belief Endurance* are all implied by orthodox thinking about various aspects of the possession conditions for true beliefs (at a time or over time).[24] The remaining three assumptions require further elaboration. *Subvenience* articulates a necessary condition for belief retention, namely, that the continued existence of a belief depends on the continued existence of an (that is, some or other) appropriate subvening base.[25]

---

[23]*Compositionality* and *Belief Endurance* together provide a plausible path to belief-alteration: the representation of a component of *P* might change, which changes the belief in question (it is no longer a belief that *P*). We acknowledge the possibility, however, that the content of a belief might change in some other, more holistic manner. The second argument below, put specifically in terms that presuppose a combinatorial semantics for belief states, might well be recast in a way not so focused on the changes in the referents of "sub-sentential" components, although we make no attempt to work out such an alternative formulation here.

[24]Note, moreover, that our use of 'proposition' is meant to be neutral with respect to the metaphysics of propositions; our assumptions align with what Cappelen and Hawthorne (2009) call 'The Simple View' according to which propositions are taken at least to play certain functional roles characteristically attributed to propositions – *viz.* as the primary bearer of truth values, the objects of agreement or disagreement, etc.

[25]The subvenience assumption, it should be clear, is stated at a level of generality such that it is applicable to *occurrent beliefs* and *dispositional beliefs* alike. Even if one is not occurrently believing a proposition *P*, one may nonetheless dispositionally believe *P*, provided one is disposed to affirm *P* and has the relevant content stored in memory (which allows for a variety of supervenience bases). If the memory trace is lost, so is the dispositional belief. For further discussion of this distinction, see Schwitzgebel (2015, §2.1). Note that dispositional beliefs are importantly different from *dispositions to believe.* The content apposite to the former must be at least stored in memory for the dispositional belief to persist. In contrast, a subject may have dispositions to believe (but not dispositional beliefs) contents she has never explicitly represented. For the

Ongoing state and *Maintenance*, which one of us has argued for in previous work, should not be controversial. Nonetheless, some contemporary writers on epistemic value obscure the point. In particular, consider again Zagzebski's analogy with coffee production. The idea in play was that a good-tasting cup of coffee takes on no additional gustatory value simply in virtue of its being the product of a reliable coffee machine.[26] From this point, we are invited to conclude – by parity of reasoning – that it is unclear how a true belief would become additionally epistemically valuable if turns out that the true belief was not only true, but also the product of a reliable belief-forming process.

Consider a different analogy. The project of maintaining a pleasant home is rather unlike the project of making a cup of coffee. The property of being well-maintained – even though it contributes instrumentally to the home's being a pleasant home – is not a property that could be 'swamped' by the value of an already pleasant flat. After all, *if the flat is going to continue to be pleasant, it will have to go on being well-maintained* as the home continues to persist; and accordingly, the property of being well-maintained can continue to confer value to the home indefinitely (2013, 256).

What separates beliefs that are candidates for knowledge (i.e., ones which are justified) from mere true beliefs is precisely what separates (by analogy) more generally ongoing states that are positively evaluable from those that are not; the former are sustained through good maintenance that the latter are not.

We propose that oftentimes—even if not always—subpersonal, justification-conferring processes, of the very sort that can help to explain why true beliefs arrived at via cognitively integrated virtues are more valuable than otherwise, underlie the continuing *existence* of a belief.[27] In many cases, for example, a belief is not held onto absent certain subpersonal justificatory processes. The belief no longer exists – either by the elimination of its subvening states or by the alteration of its content so as to make it a different belief state – if it is not effectively justified in an ongoing manner. In fact, in a wide range of cases, it is highly unlikely that a human subject has – for very long, anyway – a merely true belief. After all, if it is a true, belief-like state *but does not become cemented by subpersonal processing that is also justification-conferring*, it is oftentimes eliminated, either by a change in content or by the elimination altogether of its subpersonal basis.

Argument 1

---

canonical presentation of this distinction, see Audi (1994).

[26]The thought seems to be that, once one has a good-tasting cup of coffee, it remains good, or to the extent that it does not, the degradation of flavor likely results from, for example, chemical interactions with the surrounding gas molecules, independent of the production process; the goodness of the flavor of the two cups of coffee is, by hypothesis, a function of the appearance of the same chemical profile in them when the process that created them terminates, which screens off, from the differing processes of production, the later chemical interactions that lead to the degradation of flavor.

[27]Note that such justification-conferring subpersonal processes may allow for the subject to be epistemically responsive (e.g., to potential signs of unreliability) even if the subject never becomes consciously aware of any signs of unreliability. These are points that have been echoed in the literature on predictive processing and the Bayesian brain (e.g., Clark 2015). For further development of this point, see also.

A1. Premise 1. In order that an initially formed belief-like state be maintained long enough for it to become a full-fledged belief state, its subpersonal realizer(s) must, in many cases, be integrated into the cognitive system.

A1. Premise 2. In order that a belief endure over a significant period of time, its subpersonal realizer(s) must continue to survive integration-related subpersonal routines.

A1. Premise 3. The kinds of integration referred to in the two preceding premises contribute to the belief's justificatory status in ways that have no structural parallel at the personal level.

A1 Premise 4. In contrast, a belief-like state that is merely true is typically not integrated into the cognitive system or is not maintained after formation, and thus is likely to be eliminated.

A1 Premise 5. On the assumption that being true provides some noninstrumental value, a justified true belief has more of it than a nonexistent belief (or a belief-like state that is likely to be eliminated after a brief existence).

A1 Premise 6. Given A1 Premise 3, the account of this difference in value is distinctively subpersonal.

Therefore, in the cases in question, there is a straightforward—even if not traditionally explored—sense in which justified true beliefs are more noninstrumentally valuable than what would be the relevant merely true beliefs; and the account of this noninstrumental value is distinctively subpersonal.

It is worth making three points of clarification and elaboration. Firstly, note that scope of the first premise—viz., 'in many cases'. The premise does not have an unrestricted scope; and further scope qualifications feature elsewhere in the argument. This restriction would be problematic if it were an *ex ante* criterion of adequacy for any solution to the swamping problem that it must traffic in premises with unrestricted scopes. It is hardly clear that this is the case. As Duncan Pritchard (2013, 12) puts it, one viable way of addressing the problem is by demonstrating that 'in general and all other things being equal, we desire to be knowers as opposed to being agents who have mostly true beliefs but lack knowledge (or, worse, have mostly false beliefs)'. Argument 1 offers a vindicatory reply to the swamping problem that falls within these lines.

Secondly, we should make clear what sort of belief-cementing and belief-maintaining cognitive processes we have in mind. Central to our conception of such processes are those described in the final portion of the preceding section, pertaining to memory consolidation. But, to add further depth to the discussion, consider the process of checking for consistency. One might imagine this would occur consciously, via the conscious contemplation of the relation between one's newly acquired belief (or belief-like state) and the rest of one's beliefs. One concentrates on the content of the belief and the contents of one's other beliefs and checks the set for consistency or other, perhaps more robust, coherence-related relations. To the extent that one has justification for one's existing beliefs, consistency or

coherence with them provides justification for a newly formed belief (or belief-like state).

But, there's good reason to think that humans do not implement anything in the vicinity of this personal-level ideal. According to Christopher Cherniak's calculations, for example, the combinatorial explosion entailed by any effort to explicitly check for consistency would sink any such effort (Cherniak 1986). The subpersonal cognitive system instead uses all manner of computing tricks, typically beyond the ken of the conscious mind, to try to maintain consistency, without depending on deliberate, conscious, serial, personal-level calculation. Again, some such processes occur during slow-wave sleep, as part of a process of reactivating and strengthening representations of facts. In the cases in question, the intrinsic character that confers on a belief the power to justify is not created by or conferred on the belief by any personal-level process. There's no personal-level process to which a theorist might appeal. The only available account of the sources of this justificatory positioning lies at the subpersonal level.

As a result of this process, the beliefs in question will not only be justified, but will also be in a position to justify other beliefs, a position that is explained by subpersonal processing. Moreover, this process puts a belief in a position to be further justified by other personal-level states. Your observation of a new dog in the neighborhood might undergo consolidation and integration in a way that creates justificatory power for your resulting belief B. But, it also situates B in a collection of beliefs, vis-à-vis other beliefs about dogs, pets, ownership, etc., such that those other beliefs are more likely to maintain or increase justification for B, when appropriate. This results partly from a declarative memory's integration into existing knowledge-schemas, as happens during the consolidation (and reconsolidation) process described by Dudai et al. (2015).

Thirdly, we have used the language of being 'belief-like' to refer to states that appear in the early stages of belief-formation and that either are or might well become beliefs. We do not insist on this terminology, but we do hold that for many initially acquired or formed states of the belief-like, information-encoding sort, for that state to become a well-functioning belief (a 'full-fledged belief' one might say), it must be integrated into one's cognitive system. Furthermore, this process of integration is justification conferring, because such properties as consistency and broad coherence are justification conferring.[28]

We do not claim that contemporary cognitive science has yielded a complete and well-confirmed theory of the subconscious processes at issue; but it is highly plausible that any promising heuristic owes its efficacy to subpersonal processing. Consider consistency again. Imagine that a manageable number of randomly selected belief-realizers are activated at various times and, at each time, the active set is subject to a manageable process of consistency checking. How does one select beliefs (or their subvening structures) randomly? It boggles the mind, unless one allows some kind of fast, automatic search process, say, the selective activation of a subset of one's beliefs by the operation of an algorithm that samples from codings of them.[29]

---

[28]See, for discussion of the value of such broad coherence, Sosa (1997).

[29]In the field of artificial intelligence, it is relatively common to exploit, in various ways, the strategy of

An objector might insist that merely true beliefs possess value indistinguishable from that possessed by justified true beliefs. But as we've suggested, there are good empirical grounds to hold in many cases at least, such a merely true belief is likely to crumble and vanish quickly, for what are, from the standpoint of the personal level, inexplicable reasons. To be clear, our concern in this regard is not that a subject will be easily swayed by reasons (good or bad) to give up a merely true belief – because, as it's sometimes claimed, a merely true belief isn't, to use the Socratic metaphor, "tied down" at the personal level. Rather, very often—again, even if not always—the belief cannot even be the sort of stable thing that enters into reasons-sensitive relations until it is justified and even then, often, only insofar as it *continues* to be justified at the subpersonal level.[30]

Is the value in question noninstrumental? In some sense, it clearly is. If merely true beliefs are unstable and prone to go out of existence before they even become full-fledged belief-states – compare here with Socrates' Statues of Daedalus in the *Meno*[31] –integrated into the subject's psychology, then whatever noninstrumental value merely true beliefs have (or would have, in virtue of being true), justified true beliefs have *more* of it, if only because they last longer. To describe the comparison dramatically, the choice, at least in some cases, does not seem to be between a mere true belief and a justified true belief; it is between no belief at all and a justified true belief. Clearly the latter state of the subject is more valuable, if there's anything valuable at all about having true beliefs; any positive amount is greater than zero. In cases in which a merely true belief exists but only for a brief time, its justified, longer-lived counterpart is of greater value. The latter accrues, by virtue of its existing as a true belief for a longer period, a greater amount of whatever sort of value truth confers.

Another potential objection holds that the sorts of consolidation processes we have in mind do not confer justification on beliefs, even if such processes enhance internal consistency and coherence in the subject's overall cognitive profile. After all, some such processes serve merely to increase the coherence and maintain the consistency of what to most of us would seem to be a ridiculous set of beliefs (pertaining, say, to the subject's conviction that alien abduction has occurred). Presumably, though, such collections of beliefs (about alien abduction and the like) are thought to be ridiculous because they are false, which pushes them outside both sets of beliefs that we mean to be comparing – merely true beliefs and

---

sampling values and inferring from the properties of the sampled set something about the properties of a larger set of data or portion of the world not directly accessible to the agent. (See Russell and Norvig 2011, sections 14.5 and 15.5.3, which discuss approximate inference in Bayesian networks.) To be clear, the current point is not that the human brain is using, for instance, Markov chain Monte Carlo simulations to maintain epistemic hygiene, only the weaker point that current work in AI at least provides some clue to the sort of strategy that might be used by the brain to maintain epistemic hygiene, given that such hygiene is generally not maintained via conscious reflection on our belief and evidence sets.

[30]This conclusion comports well with Ernest Sosa's (1991) account of the value of what he calls reflective knowledge, a value that derives from a true belief's situatedness in a broadly coherent network of other true beliefs.

[31]These statues lacked a certain value, in and of themselves, given their disposition to run away if not tethered down. The value of the statues is realised only in the presence of appropriate tethering.

justified true beliefs. Thus, we can fairly set this concern aside.

We consider now a second argument, one that may be of special interest to externalist epistemologists.

Argument 2

A2. Premise 1. In many cases, subpersonal processes mediate the initial fixation of the content of a mental representation and also sustain the relations that keep its representational value fixed (by getting the mental representation into or keeping it in the content-determining relation to property, kind, or individual represented).[32]

A2. Premise 2. Such relations often involve diagnostic relations among internal representations and as such are, loosely speaking, inferential; schematically, mental representation 'a' tracks As because (i) As reliably exhibit feature B, (ii) mental representation 'b' is causally sensitive to the presence of B, and (iii) the activation of 'b' tends to cause the activation of 'a'.

A2. Intermediate conclusion 1. Subpersonal processes partly causally determine the identity of proposition believed, by determining some of the elements of the proposition believed (by Premises 1 and 2 and *Compositionality*).

A2. Intermediate conclusion 2. In some cases, subpersonal cognitive processes maintain a belief (or belief-like state), keeping it in existence by grounding the tracking relations that determine the semantic content of components of the representational structure that picks out the proposition the belief in question is an attitude toward (a change in which would eliminate that belief – per *Subvenience and Belief Endurance*).

A2. Premise 3. The internal causal relations between subpersonal states, which relations support the relevant tracking capacities, contribute to the belief's (externalist) justificatory status in ways that have no structural parallel at the personal level.

A2 Premise 4. In contrast, a merely true belief is likely to go out of existence (for if it is not externalistically justified[33] in the manner described, then it is not likely to enter into the diagnostic relations that fix and maintain its content).

Therefore, for reasons parallel to those given in connection with Argument 1, a justified true belief is, at least in many cases, more noninstrumentally valuable than a merely true belief.

In support of A2, Premise 2, we note that the relations in question diagnose the presence of the individuals, kinds, or properties represented by the subpersonal units on which

---

[32]Such diagnostic or mediating relations need not be – in fact, in most cases are not – definitional relations or expressions of necessary and sufficient conditions. On non-defining sustaining mechanisms that help to fix and maintain the content of mental representations, see Fodor (1987, 121), Margolis (1998), and Cowie (1999).

[33]Justified because, the instantiation or presence of B is, by assumption, positively diagnostic of the presence of an A. The processes at issue would provide internalist justification only if the connection between 'b' and 'a' is justificatory and internally accessible, which, given our emphasis on subpersonal processes, seems unlikely.

the belief content supervenes and that this "diagnosing" relation is generally one that is epistemically valuable. Consider a case in which one recognizes the presence of one's pet dog by slight differences in gait or scent, differences that one has difficulty articulating or bringing clearly to consciousness. In such a case, it is because of the association between subpersonal representations that one is plausibly justified in believing that Fido is in the house. Moreover, Bayesian models of cognitive processing predominate at the subpersonal level, which at least *prima facie* involve a justification or confirmation-relation in response to the environment (see Clark 2015 for more on one important class of Bayesian modeling).

Of course, accessibility internalists (Chisholm 1988, BonJour 1985) will likely express doubts about the appeal to subpersonal processes. They are almost sure to claim that subpersonal processes have nothing to do with justification, that consciousness must have direct access to anything that counts as justification and that subjects do not have conscious access to the processing in question.

Consider, however, the following possibility. Imagine that various subjects hold various beliefs that they take to be self-justifying or justified *a priori* (though fallibly so) – most importantly, justified in the absence of personal-level justifying relations. Now imagine that a pattern emerges: a significant subset of these beliefs turns out to be false, while the remaining ones turn out to be true, and there is a principled distinction between the subpersonal processes that lead to the formation and maintenance of beliefs in the former case, on the one hand, and the subpersonal processes that lead to the formation and maintenance of beliefs in the latter case, on the other hand. Focused on the personal level alone, the accessibility internalist can appeal to no structure or process that would explain the difference between the two kinds of cases; subjects report equal levels of certainty attached to both kinds of beliefs, and they report not having based them in any way on inferences. In this case, the internalist should accept that a mismatch argument establishes a role for the subpersonal in the theory of justification. We maintain that internalists should be similarly moved by mismatches of the sort we have discussed. In many cases, the processes that validate the justificatory status of states accessible to consciousness are subpersonal and lack any personal-level analogue.

Before closing, it will be helpful to register four summary points about the conclusions we have reached and to situate them in the context of the initial value of knowledge debate with which we began. Firstly, and with reference to the taxonomy introduced in the Introduction – including *validationism, fatalism* and *revisionism* – we take ourselves to have explored two importantly distinct strands of validationism, both of which involve novel recourse to the subpersonal level, and which have been hitherto ignored. The first strand of validationist response showed how two leading attempts to defend validationist strategies—developed by Olsson and Greco, respectively—are (i) inadequate as responses to the swamping problem when articulated exclusively at the personal-level; and (ii) comparatively more promising when paired with additional subpersonal-level explanations. That is to say, on this first strand of validationist response, we showed how Olson and Greco, respectively, *needed* to appeal to various ways in which subpersonal processes can be value-conferring in order to give us viable accounts of the value of knowledge *on their own favored terms*,

respectively. If this is right, then we've shown an important respect in which the role of subpersonal processes has been overlooked and has importance in epistemic axiology.

The second strand of validationist response we offered in Section Four was comparatively more ambitious. According to this strategy of response, the pretheoretical insight that the value of knowledge exceeds that of mere true opinion can in principle be vindicated *exclusively* at the subpersonal level of description. To be clear, we maintain that our case for the philosophical import of the subpersonal in epistemic axiology does not actually require this second strand of validationist response; the first strand would suffice. We have, however, attempted to show how even this stronger strand of validationist strategy has much to recommend it. In doing so, we offered two connected arguments, each of which presented a novel way in which we envisage a subpersonal response to the swamping problem being developed.

Our second and third summary comments concern the nature of this more ambitious strategy-type sketched in Section Four. The second point is that Argument 1 offers a response to the swamping problem that is in principle epistemologically ecumenical, in the sense that it by and large avoids contentious substantive epistemological assumptions. Argument 2 draws from similar insights as Argument 1, though will naturally be more attractive to externalists. This is important because, even for those epistemologists who are disinclined toward epistemic externalism, Argument 1 offers one straightforward avenue for vindicating, exclusively at the subpersonal level, the value of knowledge over mere true belief.

The third summary comment, also concerning the more ambitious forms of validationism sketched in Section Four, has to do with the *criterion of adequacy* we should expect for any viable validationist response. If, as Duncan Pritchard (2013, 12) puts it, one legitimate way of articulating the insight about the value of knowledge that must be vindicated is that 'in general and all other things being equal, we desire to be knowers as opposed to being agents who have mostly true beliefs but lack knowledge' then our more ambitious arguments, Argument 1 and Argument 2, are clearly legitimate responses. That said, we recognize that there are ways of construing a criterion of adequacy for a validationist response with reference to which Arguments 1 and Arguments 2 would not be legitimate. For instance, if such a response requires vindicating the (we believe, unduly strong) pretheoretical insight that all *possible* items of knowledge have a value that exceeds the value of comparative true belief tokens, then Arguments 1 and 2 look less promising. However—as Kvanvig (2003) himself has argued at length—it's not at all obvious that there is, at the end of the day, *any* way of defending a validationist response if such a strong modal criterion of adequacy is what's assumed. At any rate, it will suffice for our purposes to register that our more ambitious arguments in Section Four show two new ways of defending entirely subpersonal validationist responses to the value problem which will be regarded as legitimate with reference to the kind of criterion of adequacy Pritchard has in mind, and which isn't predicated upon what are perhaps overoptimistic pretheoretical assumptions about what it *is* that should be validated.

Fourth, we admit, even emphasize, our limited understanding of the nature of the subpersonal processes in question. But, we take ourselves to have provided sufficient reason to be enthusiastic about the present approach as a way to make progress on questions about epistemic value, generally speaking, and the swamping problem, in particular.[34]

References

Audi, Robert. 1994. "Dispositional Beliefs and Dispositions to Believe." *Nous* 28 (4):419–434.

Baehr, J. 2009. "Is There a Value Problem?" In *Epistemic Value*, edited by A. Haddock, A. Millar, and D. Pritchard. Oxford: Oxford University Press.

Barbar, Sarah J., Suparna Rajaram, and Elizabeth J. Marsh. 2008. "Fact Learning: How Information Accuracy, Delay, and Repeated Testing Change Retention and Retrieval Experience." *Memory* 16, 8: 934–946.

Barnier, Amanda J., John Sutton, Celia B. Harris, and Robert A. Wilson. 2008. "A Conceptual and Empirical Framework for the Social Distribution of Cognition: The Case of Memory." *Cognitive Systems Research* 9 (1):33–51.

Bechara, A., H. Damasio, D. Tranel, and A. R. Damasio. 2005. "The Iowa Gambling Task and the Somatic Marker Hypothesis: Some Questions and Answers." *Trends in Cognitive Sciences* 9 (4):159-162-164.

Begg, Ian Maynard, Ann Anas, Suzanne Farinacci. 1992. "Dissociation of Processes in Belief: Source Recollection, Statement Familiarity, and the Illusion of Truth." *Journal of Experimental Psychology: General* 121, 4: 446–458.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge University Press.

———. 1998. *In Defense of Pure Reason*. Cambridge University Press.

Brown, Alan S., and Elizabeth J. Marsh. 2010. "Digging into Déjà Vu: Recent Research on Possible Mechanisms." In In Brian H. Ross (Ed.), *Psychology of Learning and Motivation* vol. 53, pp. 33–62. Burlington: Academic Press.

Brown, Evan, Kenneth Deffenbacher, & William Sturgill. 1977. "Memory for Faces and the Circumstances of Encounter." *Journal of Applied Psychology* 62, 3: 311–318.

Cappelen, H., and J. Hawthorne. 2009. *Relativism*

---

[34]The reader might worry that our suggestions do not address the swamping problem for justified true belief neat, absent any commitment to the way it appears or is maintained in a particular kind of physical system. In response to such concerns, one of us is inclined to defend what might seem to be a strict naturalistic line: Human investigators have no experience with, or direct cognitive access to, nonimplemented beliefs or to beliefs implemented in any other kind of materials from the materials human beliefs are implemented in; it is these states that, as a matter of fact, are the target of introspection and that causally interact with the output of introspection (including reports of intuitions about matters epistemic). Thus, far all we know, philosophers' intuitions about justified true belief are, in fact, intuitions about the one case we know directly, the human case, and bear marks of their interactions with the fine-grained nature of the human case; if so, a solution to the swamping problem for the human case is just what those intuitions cry out for. Jointly, however, we endorse a more ecumenical rejoinder: we have argued that, as they appear in at least one believing system, a significant number of justified true beliefs have more noninstrumental value than the corresponding merely true beliefs, and thus the swamping problem is solved for at least some cases.

and Monadic Truth. Oxford: Oxford University Press.

Carter, J. Adam, Andy Clark, and S. Orestis Palermos. 2018. "New Humans: Ethics, Trust and the Extended Mind." In *Extended Epistemology*. Oxford: Oxford University Press.

Chambers, Karen L., and Maria S. Zaragoza. 2001. "Intended and Unintended Effects of Explicit Warnings on Eyewitness Suggestibility: Evidence from Source Identification Tests." *Memory and Cognition* 29, 8: 1120–1129.

Cherniak, Christopher. 1986. *Minimal Rationality*. Cambridge, MA: MIT Press

Chisholm, Roderick M. 1988. "The Indispensability of Internal Justification." *Synthese* 74 (3): 285–296.

Clark, Andy. 2001. "Visual Experience and Motor Action: Are the Bonds Too Tight?" *Philosophical Review* 110 (4):495–519.

———. 2015. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Clark, Andy, and David J. Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1):7–19.

Cowie, Fiona. 1999. *What's Within: Nativism Reconsidered*. Oxford: Oxford University Press.

Davies, Martin. 2000*a*. "Interaction without Reduction: The Relationship between Personal and Sub-Personal Levels of Description." *Mind & Society* 1: 87–105.

———. 2000*b*. "Persons and Their Underpinnings." *Philosophical Explorations* 3: 43- 62.

de Brigard, Felipe. 2017. "Cognitive Systems and the Changing Brain." *Philosophical Explorations* 20 (2): 224–241.

Dechêne, Alice, Christoph Stahl, Jochim Hansen, and Michaela Wänke. 2010. "The Truth About the Truth: A Meta-Analytic Review of the Truth Effect." *Personality and Social Psychology Review* 14, 2: 238–257.

Dennett, Daniel C. 1969. *Content and Consciousness*. Routledge.

———. 1991. *Consciousness Explained*. Boston: Little, Brown and Company.

Dewhurst, Stephen A., Martin A. Conway, and Karen R. Brandt. 2009. "Tracking the R- to-K Shift: Changes in Memory Awareness across Repeated Tests." *Applied Cognitive Psychology* 23: 849–858.

Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.

Drayson, Zoe. 2012. "The Uses and Abuses of the Personal/Subpersonal Distinction." *Philosophical Perspectives* 26 (1):1–18.

———. 2014. "The Personal/Subpersonal Distinction." *Philosophy Compass* 9 (5):338–346.

Dudai, Yadin, Avi Karni, and Jan Born. 2015. "The Consolidation and Transformation of Memory." *Neuron* 88: 20–32.

Echterhoff, Gerald, William Hirst, and Walter Hussy. 2005. "How Eyewitnesses Resist Misinformation: Social Postwarnings and the Monitoring of Memory Characteristics." *Memory and Cognition* 33, 5: 770–782.

Evans, Jonathan St. B. T. and Keith Frankish (eds.). 2009. *In Two Minds: Dual Processes and Beyond*. Oxford University Press.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh. 2015. "Knowledge Does Not Protect Against Illusory Truth." *Journal of Experimental Psychology: General* 144, 5: 993–100.

Fodor, Jerry. 2007. *Psychosemantics*. Cambridge, MA: MIT Press.

Frankish, Keith. 2009. "Systems and Levels: Dual-System Theories and the Personal-Subpersonal Distinction." In Evans and

Frankish 2009, pp. 89–107.

Gawronski, Bertram, and Galen V. Bodenhausen. 2014. "The Associative–Propositional

Evaluation Model: Operating Principles and Operating Conditions of Evaluation."

In J. W. Sherman, B. Gawronski, and Y. Trope (eds.), *Dual-Process Theories of*

the Social Mind, pp. 188–203. New York: Guilford Press.

Gendler, Tamar. 2008. "Alief and Belief," *Journal of Philosophy* 105 (10): 634–663.

Geraci, Lisa, and Nancy Franklin. 2004. "The Influence of Linguistic Labels on Source- Monitoring Decisions." *Memory* 12, 5: 571–585.

Gertler, Brie. 2011. *Self-Knowledge*. New York: Routledge.

Goldman, A. 1979. What is Justified Belief? In George Pappas (ed.), *Justification and Knowledge*. Boston: D. Reidel. pp. 1-25.

Goldman, A., and E. J. Olsson. 2009. "Reliabilism and the Value of Knowledge." In *Epistemic Value*, edited by A. Haddock, A. Millar, and D. Pritchard. Oxford: Oxford University Press.

Goldstone , R. L., and T. M. Gureckis. 2009. "Collective Behavior." *Topics in Cognitive Science* 1 : 412–38.

Greco, John. 2010. *Achieving Knowledge*. Cambridge: Cambridge University Press.

Haddock, Adrian, Alan Millar, and Duncan Pritchard. 2009. *Epistemic Value*. Oxford University Press.

———. 2010. *The Nature and Value of Knowledge: Three Investigations*. Oxford University Press.

Hasher, Lynn, David Goldstein, and Thomas Toppino. 1977. "Frequency and the Conference of Referential Validity." *Journal of Verbal Learning and Verbal Behavior* 16: 107-112

Haybron, Daniel M. 2007. "Do We Know How Happy We Are? On Some

Limits of Affective Introspection and Recall." *Noûs* 41 (3): 394–428.

Heck, Richard G. 2000. "Nonconceptual Content and The 'Space of Reasons.'" *The Philosophical Review* 109 (4):483–523.

Henkel, Linda A., and Mark E. Mattson. 2011. "Reading Is Believing: The Truth Effect and Source Credibility." *Consciousness and Cognition* 20: 1705–1721.

Huemer, Michael. 2007. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74: 30–55.

Huijding, Jorg, and Peter J. de Jong. 2009. "Implicit and Explicit Attitudes toward Spiders: Sensitivity to Treatment and Predictive Value for Generalization of Treatment Effects." *Cognitive Therapy and Research* 33: 211–220.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jacoby, Larry L., Colleen Kelley, Judith Brown, and Jennifer Jasechko. 1989. "Becoming Famous Overnight: Limits on the Ability to Avoid Unconscious Influences of the Past. *Journal of Personality and Social Psychology* 56, 3: 326–338.

Jacoby, Larry L., Vera Woloshyn, and Colleen Kelley. 1989. "Becoming Famous Without Being Recognized: Unconscious Influences of Memory Produced by Dividing Attention." *Journal of Experimental Psychology: General* 118, 2: 115–125.

Johnson, Marcia K., Shahin Hashtroudi, and D. Stephen Lindsay. 1993. "Source Monitoring." *Psychological Bulletin* 114, 1: 3–28.

Jones, Matt, Tim Curran, Michael C. Mozer, and Matthew H. Wilder. 2013. "Sequential Effects in Response Time Reveal Learning Mechanisms and Event Representations." *Psycho-*

logical Review 120 (3):628.

Klein, Colin. 2010. "Critical Notice: Cognitive Systems and the Extended Mind by Robert Rupert" *The Journal of Mind and Behavior* 31 (3&4): 253–264.

Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

———. 2008. "Pointless Truth." *Midwest Studies in Philosophy* 32 (1):199–212.

———. 2010. "'The Swamping Problem Redux: Pith and Gist.'" In *Social Epistemology*, edited by Adrian Haddock, Alan Millar, and Duncan Pritchard, 89–112. Oxford: Oxford University Press.

Lewis, David. 1972. "Psychophysical and Theoretical Identification." *Australasian Journal of Philosophy* 50: 249–258

Loftus, Elizabeth. 1979. *Eyewitness Testimony*. Cambridge: Harvard University Press.

Margolis, Eric. 1998. "How to Acquire a Concept." *Mind & Language* 13 (3): 347–369.

Marsh, E. J., A. N. Eslick, and L. K. Fazio. 2008. "False Memories." In H. L. Roediger, III (Ed.), *Cognitive Psychology of Memory*. Vol. [2] of *Learning and Memory: A*

Comprehensive Reference, 4 vols. (J.Byrne Editor), pp. 221–238. Oxford: Elsevier.

Mather, Mara, Eldar Shafir, and Marcia K. Johnson. 2000. "Misrememberance of Options

Past: Source Monitoring and Choice." *Psychological Science* 11, 2: 132–138.

McDowell, John. 1994*a*. "The Content of Perceptual Experience." *The Philosophical Quarterly* 44 (175):190–205.

———. 1994*b*. *Mind and World*. Cambridge, MA: Harvard University Press.

Meade, Michelle L., and Henry L. Roediger III.

2002. *Memory & Cognition* 30, 7: 995– 1009.

Menary, Richard. 2010. "The Extended Mind and Cognitive Integration." In *The Extended Mind*, edited by Richard Menary. MIT Press.

———. 2012. "Cognitive Practices and Cognitive Character." *Philosophical Explorations* 15 (2):147–164.

Moore, G. E. 1903. *Principia Ethica*. Dover Publications.

Myers-Schulz, Blake, and Eric Schwitzgebel. 2013. "Knowing That P without Believing That P." *Noûs* 47 (2):371–384.

Nisbett, Richard E., and Timothy D. Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3):231–59.

Olsson, E. J. 2007. "Reliabilism, Stability and the Value of Knowledge." *American Philosophical Quarterly* 44 (4):343–55.

Perugini, Marco. 2005. "Predictive Models of Implicit and Explicit Attitudes," *British Journal of Social Psychology* 44: 29–45.

Piccinini, Gualtiero. 2010. "How To Improve on Heterophenomenology: The Self- Measurement Methodology of First-Person Data." *Journal of Consciousness Studies* 17 (3–4): 84–106.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

Pritchard, Duncan. 2007. "Recent Work on Epistemic Value." *American Philosophical Quarterly*, 85–110.

———. 2009a. "Knowledge, Understanding and Epistemic Value." *Royal Institute of Philosophy Supplement* 64:19–43.

———. 2009b. "The Value of Knowledge." *The Harvard Review of Philosophy* 16 (1):2–19.

———. 2010. "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175

(1):133–151.

———. 2011. "What Is the Swamping Problem." In *Reasons for Belief*, edited by Andrew Reisner and Asbjorn Steglich-Petersen. Cambridge: Cambridge University Press.

———. 2013. *What Is This Thing Called Knowledge?* Routledge.

Rabinowicz, Wlodek, and Toni Ronnow-Rasmussen. 2000. "II-A Distinction in Value: Intrinsic and For Its Own Sake." *Proceedings of the Aristotelian Society* 100 (1):33–51.

Rasch, Bjorn, and Jan Born. 2013. "About Sleep's Role in Memory." *Physiological Reviews* 93 (2):681–766.

Reber, Rolf, and Norbert Schwarz. 1999. " Effects of Perceptual Fluency on Judgments of Truth." *Consciousness and Cognition* 8: 338–342.

Rey, Georges. 2001. "Physicalism and Psychology: A Plea for a Substantive Philosophy of Mind," in C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press, pp. 99–128.

Ridge, Michael. 2013. "Getting Lost on the Road to Larissa." *Noûs* 47: 181–201.

Riggs, Wayne. 2008. "The Value Turn in Epistemology." In *New Waves in Epistemology*, edited by Vincent Hendricks, 300--23. Palgrave Macmillan.

Rønnow-Rasmussen, Toni. 2011. *Personal Value*. Oxford University Press.

Rupert, Robert D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101, 389–428.

Rupert, Robert D. (2009). *Cognitive systems and the Extended Mind*. Oxford: Oxford University Press.

Rupert, Robert D. (2013). Memory, natural kinds, and cognitive extension; or, Martians don't remember, and cognitive science is not about

cognition. *Review of Philosophy and Psychology* 4: 25–47.

Russell, Stuart, and Peter Norvig. 2011. *Artificial Intelligence: A Modern Approach*. Third Edition. Upper Saddle River, NJ: Pearson.

Schwitzgebel, Eric. 2008. "The Unreliability of Naive Introspection." *Philosophical Review* 117: 245–273.

Shea, Nicholas. 2013. "Neural mechanisms of decision-making and the personal level." In K.W.M. Fulford, M. Davies, G. Graham, J. Sadler, G. Stanghellini, and T. Thornton (eds.) *Oxford Handbook of Philosophy and Psychiatry*. Oxford: Oxford University Press, pp. 1063–1082.

Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge University Press.

———. 1997. "Reflective Knowledge in the Best Circles." *The Journal of Philosophy* 94 (8):410–430.

———. 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge (Vol. 1)*. Oxford: Oxford University Press.

Sporns, Olaf, Dante R. Chialvo, Marcus Kaiser, and Claus C. Hilgetag. 2004 "Organization, Development and Function of Complex Brain Networks." *TRENDS in Cognitive Sciences* 8 (9): 418–425

Squire, Larry R., Lisa Genzel, John T. Wixted, and Richard G. Morris. 2015. "Memory Consolidation." *Cold Spring Harbor Perspectives in Biology doi:10.1101/cshperspect.a021766*

Srivastava, Vipin, Suchitra Sampath, and David J. Parker. 2014. "Overcoming Catastrophic Interference in Connectionist Networks Using Gram-Schmidt Orthogonalization." *PLoS ONE 9(9) doi:10.1371/journal.pone.0105619*

Sylvan, Kurt. 2017. "Veritism Unswamped." *Mind*. 127 (506): 381–435

Tosun, Suimeyra, Jyotsna Vaid, and Lisa Geraci. 2013. "Does Obligatory Linguistic Marking of Source of Evidence Affect Source Memory? A Turkish/English Investigation." *Journal of Memory and Language* 69: 121–134.

Tulving, Endel. 1972. "Episodic and Semantic Memory." In E. Tulving & W. Donaldson, *Organization of Memory*. Oxford: Academic Press.

Vendler, Zeno. 1957. "Verbs and Times." *The Philosophical Review*, 143–160.

Wegner, Daniel. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.

Wilson, Timothy D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Belknap Press of Harvard University Press.

Wilson, Margaret. 2002. "Six Views of Embodied Cognition," *Psychonomic Bulletin and Review* 9: 625–636.

Zagzebski, Linda. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.

———. 2003. "The Search for the Source of Epistemic Good." *Metaphilosophy* 34 (1–2):12–28. https://doi.org/10.1111/1467-9973.00257.