Unconscious Evidence

Jack Lyons

University of Arkansas

Can beliefs that are not consciously formulated serve as part of an agent’s evidence for other beliefs? A common view says no, any belief that is psychologically immediate is also epistemically immediate. I argue that some unconscious beliefs can serve as evidence, but other unconscious beliefs cannot. Person-level beliefs can serve as evidence, but subpersonal beliefs cannot. I try to clarify the nature of the personal/subpersonal distinction and to show how my proposal illuminates various epistemological problems and provides a principled framework for solving other problems.

When I look around the room, I don’t normally think about how things look, or what kinds of experiences I’m having, or whether perception is reliable in these circumstances. When you give me directions to the train station, I don’t usually run through an inductive argument about the sincerity and competence of speakers. These things that I’m not thinking about are all things that I in some sense believe, but they are not things that I consciously dwell on in forming judgments about chairs, tables, people, train stations, and the like. But what is the evidential status of these beliefs?

One traditional answer to this question holds that these sorts of beliefs do indeed serve as an essential part of the agent’s evidence for her ordinary beliefs about external objects and their properties, even when the former are not consciously and explicitly formed. A competing view holds that these beliefs, when not consciously formed, don’t figure into the agent’s evidence for the more ordinary beliefs. I will call the first position a “Cartesian view” and the second a “Reidian view.” Cartesian views include not only classical foundationalism but standard forms of coherentism as well. Modest foundationalists, reliabilists, anti-reductionists about testimony, and reformed epistemologists all tend to endorse the Reidian view.

Reidians incline toward a liberal view about epistemically basic beliefs---i.e., beliefs that don’t require evidential relations to other beliefs for their justification---and a restrictive view about what is required for one belief to serve as evidence for another belief. Reidian views are a diverse lot, and arguments for the position are often inexplicit, but one important source of support for the view seems to derive from the following principle:

**Reidian Criterion**: if a belief is psychologically immediate, then it is epistemically noninferential (i.e., if justified, basically justified)[[1]](#footnote-2)

What is mean here by ‘psychologically immediate’ is what BonJour meant by ‘cognitively spontaneous’: the belief “is not arrived at via any sort of conscious ratiocinative process, but simply occurs to me, strikes me, in a coercive manner over which I have no control” (1976, p. 291).

The most obvious instances of psychologically immediate beliefs are what Sellars (1979) called IPM judgments: introspection, perception, and memory beliefs. But, as hinted above, beliefs from testimony might satisfy the criterion as well. And once we’ve started down this road, some religious beliefs may count as basic, too. So too would a number of intuitive beliefs, not only in a strict sense of ‘intuition’ popular among philosophers, but also in the broader sense popular among psychologists, where we might might have empirical intuitions about how much a product is worth, how probable it is that a person who fits a given description is a feminist bank teller, etc. The Reidian thus counts as basic all the beliefs that the Cartesian does and then many more, since all of these are psychologically immediate.

I want to argue here that the Reidian Criterion is mistaken, although not in a way that should provide any real comfort to proponents of the Cartesian view.

I should clarify some terminological issues right away. First, as for ‘conscious’ and ‘belief’, I won’t say much about either; these are fraught terms, but none of my discussion hinges on a very specific understanding of either. Many of the “beliefs” I will be concerned with, for example, are better labeled “assumptions,” especially since the latter term suggests something one relies on without perhaps consciously formulating, but I won’t deny that they are beliefs, as I don’t want to decide on semantic or folk psychological grounds whether they can play a doxastic role for epistemic purposes.

The term ‘evidence’ has a narrow and specialized meaning in epistemology. Evidence is not just any factor that conduces to justification but is one specific *kind* of factor. One’s evidence for a belief is one’s reasons, or grounds for that belief; it it that on which the belief is based. Evidence is thus to be contrasted with coherence, reliability, assertoric force, clarity and distinctness, or whatever other factors one might think justification supervenes on. Coherence confers justification (let’s suppose) and evidence confers justification, but not in the same way.[[2]](#footnote-3)

In light of this distinction, beliefs might be relevant to the justification of other beliefs in both evidential and nonevidential ways. Imagine an epistemology in the spirit of Foley (1987), which holds that whether *e* is evidence for me that *h*, depends on my deep-seated convictions about what counts as evidence for what. These convictions need not serve as additional pieces of evidence; they might, instead, serve a metaevidential role of constitutively determining evidential relations. From a certain kind of internalist perspective, something like this might offer an attractive solution to Lewis Carroll’s (1895) regress problem: it’s my belief that *modus ponens* is valid that makes *modus ponens* valid for me,[[3]](#footnote-4) but this belief doesn’t serve as evidence and thus doesn’t constitute another premise and doesn’t generate a vicious regress. Instead, the belief serves as part of the supervenience base by which evidential relations are determined.

Finally, I want to clarify my understanding of the claim that a belief is epistemically noninferential (equivalently, that it is epistemically basic). I will say that a belief is epistemically basic iff its justification does not depend on *evidential* relations to other beliefs. This understanding of basicality would allow a basic belief’s justification to depend on evidential relations to things that aren’t beliefs (e.g., sense experiences) and would allow a basic belief’s justification to depend on nonevidential relations to other beliefs, as in the Foley-esque scenario just described.

1. **Doxastic Justification and a Causal Requirement**

My concern here is with doxastic justification, rather than with propositional justification, i.e., not with whether *p* is the right thing for an agent to believe, given her current epistemic position, but whether her currently believing that *p* is epistemically justified. Given this, I think the following causal requirement on evidence —in the sense relevant to doxastic justification—should not be very controversial:

**(CR)** If *e* is part of *S*’s evidence for believing that *p*, then *e* is causally implicated in *S*’s believing that *p*.

Causal implication is intended to be broad and inclusive. The requirement demands only that *e* make some causal contribution to *S*’s belief, not, for example, that the belief wouldn’t have existed without *e*, etc.

CR does require more for (doxastic-justification-providing) evidence than that it be something an agent could cite if challenged. Such items surely contribute to propositional justification, but not, it seems, to doxastic justification. It is common, in fact, to define doxastic justification in terms of propositional justification plus a causal connection (e.g., Conee and Feldman 2004).

CR does not beg the question against a Cartesian epistemology, not even of a coherentist variety. Beliefs are presumably encoded in synaptic connections among neurons, and existing synaptic connections surely have a causal influence on new beliefs, so there is no in-principle obstacle to *my whole belief set* being causally implicated in any given judgment. It is, of course, an empirical issue how much of one’s belief set is causally implicated in an agent’s having any given token belief, and it may not be empirically plausible that the whole belief set is implicated each time, but my point here is merely that there’s no in-principle obstacle, that CR doesn’t beg the question against the Cartesian view.

**2. Against the Reidian Criterion**

One important way that evidence differs from other factors relevant to justification is that the epistemic status of evidence (when it has an epistemic status) matters to its ability to confer justification. Unjustified beliefs, for example, can’t serve as justifying evidence for another belief. This commonplace observation leads to an important principle linking evidence and justification:

**(EJ)** Nothing that is unjustified can be justifying evidence for a belief.

EJ does not, of course, deny that a belief can be based on something that is unjustified; but when it is, being thus based does not contribute to that belief’s justification. Note that EJ is substantially weaker than the claim that only justified things can serve as justifying evidence. Lots of things—most notably experiences—are neither justified nor *un*justified. And of course, many epistemologists hold that experiences can serve as justifying evidence for beliefs, and that they don’t need to be justified in order to do so. That claim does not conflict with EJ.[[4]](#footnote-5)

If experiences can indeed justify beliefs and can do so without being themselves justified, then this is surely only because they are capable of occupying a middle ground between justified and unjustified. Like rocks, or pains, they’re neither (epistemically) justified nor unjustified. Beliefs, on the other hand, don’t seem capable of falling into this middle space; they are either justified or unjustified. Thus, EJ implies that a belief can only serve as justifying evidence if it’s itself justified:

**(EJB)** The only beliefs that can be justifying evidence for a belief are justified beliefs.

This principle will be an important aid in determining whether a given belief is playing an evidential role or not.[[5]](#footnote-6) If a belief is playing an evidential role, then that belief’s justificational status affects its evidential contribution to the justificandum belief. Conversely, if the justificational status of a belief doesn’t have any effect on the evidential/justificatory status of some justificandum belief, then it’s not providing evidential support.

Consider then, the following kind of case (Senor 2009): Looking at the sky, you think, “what a beautiful sunset!” You can’t tell the difference between sunrises and sunsets merely by looking at them, but you’re aware that it’s evening, rather than morning. Nevertheless, the belief that it’s a beautiful sunset is psychologically immediate; in particular, you don’t consciously token the belief that it’s evening. Presumably the sunset belief depends, both causally and epistemically, on the belief that it’s evening. That’s part of *why* you think it’s a sunset, rather than a sunrise. And if you weren’t justified in believing it was evening, you wouldn’t be justified in believing it was a sunset.

The unconscious reasoning can be more complex, as well (Lyons 2014):When I’ve left the top down on my Jeep, but I’m in the house, a particular sound on the roof makes me immediately form the alarmed belief that the seats are getting wet. This belief is psychologically immediate; in particular, it is not accompanied by any conscious tokenings of the beliefs that the top is down; that it’s now raining; that if it’s raining and the top is down, the seats are getting wet, etc. Yet the belief that the seats are getting wet causally depends on them; I don’t experience that immediate sense of alarm when I haven’t had the top down, or if I know the car is in the garage. And there seems to be an epistemic dependence as well: I wouldn’t be justified in thinking the seats were getting wet if I weren’t justified in thinking that the top was down, etc.

Both cases involve beliefs that are psychologically immediate although partly causally dependent on other, unconscious, beliefs. The epistemic status of the psychologically immediate beliefs, furthermore, seems to be linked to the epistemic status of these unconscious beliefs. The claim that these unconscious beliefs were serving as justifying evidence for the psychologically immediate beliefs would, by EJB, explain this linkage. Therefore, it’s plausibly the case that these unconscious beliefs are serving as *evidence* for the other beliefs, despite the psychological immediacy of the latter.

**3. Complications and Restrictions**

Suppose I’m right that the two examples just offered involve cases where the epistemic status of a psychologically immediate belief does depend on the epistemic status of some other belief. By itself this is not a conclusive reason to reject the Reidian Criterion. This is because epistemic dependence doesn’t entail evidential dependence.

Let’s say that the belief that *p* **epistemically depends** on the belief that *q* iff the belief that *p* wouldn’t be justified unless the belief that *q* were justified. And let’s say that the belief that *p* **evidentially depends** on the belief that *q* iff the belief that *p* is based on the belief that *q*, i.e., the belief that *q* is part of one’s evidence that *p*.

Suppose, controversially, that my belief that there’s a chair in front of me epistemically depends on my belief that perception is reliable, or that I’m having such-and-such a visual experience, or that the object in front of me has such-and-such low level perceptible properties. This means only that my belief that there’s a chair wouldn’t be justified if these other beliefs weren’t justified. Would this establish the Cartesian claim that perceptual beliefs *evidentially* depend on beliefs about reliability (etc.)? No, because epistemic dependence doesn’t imply evidential dependence. It is also true, for example, that in my current situation, I wouldn’t be justified in

(a) the belief that there is a chair in front of me

if I weren’t justified in

(b) the belief that there is a physical object in front of me.

If I weren’t justified in (b), it would be because I had a reason to distrust my senses, and that would serve as a defeater for (a). But surely this doesn’t show that (a) is evidentially dependent on (b). On the contrary, if there’s any evidential dependence here, it’s surely that (b) is evidentially dependent on (a). Epistemic dependence, as defined above, is a species of counterfactual dependence, which is to say that it only requires a certain counterfactual to be true and doesn’t require any *real* dependence at all. *p* could counterfactually depend on *q* while *q* counterfactually depends on *p*; similarly for the special case of epistemic dependence.

The current example illustrates epistemic dependence without evidential dependence, in part because it illustrates another species of counterfactual dependence without causal dependence. It’s probably true that I wouldn’t believe there was a chair there if I didn’t believe there was a physical object there—the nearest possible world in which I don’t in some sense believe there’s a physical object there is one where I’m suspending belief on such a wide range of things that I don’t believe there’s a chair there either—but this doesn’t mean that (a) is causally dependent on (b). In the earlier sunset and Jeep cases, we do have causal dependence, and this causal dependence partly explains the epistemic dependence. That’s why we should conclude that the epistemic dependence in these cases results from evidential dependence.

It’s important to point out, however, that causal dependence, by itself, is little indication of evidential dependence. Those “beliefs,” or assumptions,[[6]](#footnote-7) that are embedded in subpersonal modules are causally implicated in the formation of perceptual and other judgments, but their epistemic status is irrelevant to the epistemic status of those judgments.[[7]](#footnote-8) The normal human visual system, for example, operates on several assumptions, including that nearby objects are lit from above, and that retinally adjacent points are probably roughly equidistant from the perceiver. An ordinary agent, who isn’t versed in the principles of perceptual psychology, might well have very good reason to doubt such claims about light sources or retinal adjacency, rendering these beliefs or assumptions unjustified. But this wouldn’t threaten the justification of her perceptual beliefs, at least not so long as she was unaware of how such assumptions are causally relevant to things looking the way they do. But if a belief/assumption’s being unjustified doesn’t detract from the justification of beliefs that causally depend on it, then the assumption isn’t serving as *evidence* for those beliefs. It might still play an epistemic role, by conducing to reliability, etc., but it is not a specifically evidential role.[[8]](#footnote-9)

**4. Personal and Subpersonal Beliefs**

I have been arguing that some unconscious beliefs/assumptions have evidential import and that some do not. What’s the difference? My proposal is that the person-level assumptions that are causally implicated in the formation of occurrent judgments are part of the evidential basis of these judgments; these judgments are then nonbasic; they’re epistemically inferential. Subpersonal assumptions, on the other hand, are not evidentially relevant, even to those judgments they causally contribute to. In the rest of this section I will try to defend and clarify this proposal.

Most obviously and most fundamentally, the personal/subpersonal distinction is a distinction between *states of the agent* and states of something contained in the agent.[[9]](#footnote-10) My toddler daughter’s visual system works normally; it assumes that retinally adjacent points are probably roughly equidistant. This isn’t something *my daughter* assumes; she lacks the relevant concepts, and even she had them, she wouldn’t immediately be inclined to think the principle was true. I will soon try to say a bit more about which states are attributable to the agent and which are not, but I think the distinction is tolerably clear already. Insofar as we understand statements of the form ’*S* believes that *p*’, we understand, to that degree, the personal end of the personal/subpersonal distinction. And to the extent that we understand claims of the form ‘the visual system assumes that *p*’, we understand the subpersonal end of the distinction. If *S* believes that *p*, then the belief that *p* is a person-level belief of *S*. If *S* doesn’t believe that *p*, but some processing system within *S* contains the assumption that *p*, then the assumption that *p* is a subpersonal assumption. It should be clear that I don’t mean this distinction to reproduce the conscious/unconscious distinction, or probably even the conscious-*in-principle*/not distinction (although it’s hard to know, as the notion of consciousness in principle is less clear than the notion of attributability to the agent). All subpersonal states are unconscious; not all unconscious states (and maybe not all in principle unconscious states) are subpersonal.

Paradigmatic instances of subpersonal states and processes are those embedded in modular systems. Perceptual systems exhibit a number of interesting features first pointed out by Fodor (1983): they are, at least to an interesting degree, fast, automatic, innately specified, introspectively opaque, and—relatively—informationally encapsulated. The so-called System 1 systems (Stanovich & West 2000, Kahneman 2011), which are responsible at least for our empirical intuitions and likely our a priori intuitions as well, have similar properties. I will consider all these systems to be at least weakly modular.[[10]](#footnote-11) Weak modularity only requires the aforementioned features *to an interesting degree*: I doubt that perception is fully encapsulated, though clearly there are serious limits on the perceptual effects of central beliefs and goals; and surely they are not fully innately specified, at least not if that precludes perceptual learning. In fact, another conspicuous and probably essential feature of modular systems is almost never discussed, probably because of Fodor’s emphasis on innateness: modular systems are subject to an unusual and distinctive form of learning. Perceptual learning, and the learning of other modular systems, is a slow, gradual, data-driven process that requires large numbers of exposures to the relevant stimuli,[[11]](#footnote-12) that operates relatively independently of the agent’s beliefs or expectations (except insofar as these determine the allocation of attention), and whose results are specifically tied to the training conditions.[[12]](#footnote-13) Person-level learning, on the other hand, is typically fast and flexible: I only had to hear once that David Bowie died, and now I know it; the knowledge I ended up with is the same, whether I came to it by hearing it, reading visual text, or reading Braille.

There are several reasons to deny that intramodular assumptions are beliefs *of the agent*. First, having the assumption that *p* buried away in a module is far from sufficient for the agent’s having the belief that *p* (*p* might be about zero-crossings, continuity constraints, etc.). In general, the automaticity of modular systems implies that all the assumptions they contain are nonconceptual, in the “state” sense (Heck 2000): the agent needn’t have the concepts required to specify the content of some assumption, in order for the module to genuinely embody that assumption. Second, an agent’s holding a belief that conflicts with some intramodular assumption has no tendency to produce cognitive dissonance or other psychological conflict. When *I* believe two things that patently conflict, on the other hand, this tends to produce cognitive dissonance. Third, an agent’s beliefs are not *epistemically* threatened by conflicting intramodular beliefs: my belief that lighting conditions are abnormal here isn’t defeated by the conflicting assumptions of my visual system. Nor are intramodular beliefs epistemically threatened by the agent’s beliefs. (Intramodular beliefs presumably don’t have an epistemic status, but if they do, it’s not a status that’s affected by the agent’s beliefs.) But when *a person* has two beliefs that patently conflict, at least one of the beliefs suffers epistemically. Fourth, attributing the intramodular beliefs to the agent would require, where the agent consciously believes the content of that intramodular belief (e.g., that things are lit from above), attributing to the agent two beliefs with the same content, one of which is causally relevant to some judgment, and one that isn’t. This seems awkward. Fifth, it’s sometimes held that it’s impossible for someone to believe *p* and not-*p* at the same time. If so, it’s yet another reason to deny that intramodular beliefs are beliefs of the agent, for a person can clearly believe *p* while her modular systems believe not-*p*.

I’m committing to the claim that an assumption is evidentially relevant just in case it is person-level and causally relevant, and to the claim that intramodular assumptions are subpersonal. But I want to enter some more speculative proposals about the personal/subpersonal distinction as well.

Person-level beliefs, beliefs of the agent, form an (imperfectly) integrated whole; adding a new belief causes the agent to make rationally mandated compensatory changes elsewhere. Even when an agent’s beliefs persist in conflicting with each other, this normally produces cognitive dissonance and a motivation to revise, seek out particular kinds of evidence, defend, selectively interpret, confabulate, or otherwise minimize the conflict. The intramodular beliefs aren’t part of this integrated system; they can comfortably coexist alongside blatantly contradictory extramodular beliefs. Encapsulation and opacity create a kind of firewall that makes that comfort possible. Encapsulation is why my person-level assumption that things are lit from below affects neither my visual system’s assumption that things are lit from above nor the percepts that result from it. Opacity is why my visual system’s assumption doesn’t influence my contrary assumption. Intramodular assumptions are in a sense, therefore, rationally isolated from their extramodular counterparts. Person-level beliefs are often labile: a change in evidence—or change in mood—can result in a change in beliefs. Intramodular beliefs are nonlabile, alterable only by the arduous, gradual learning process described above.

I don’t want to overstate the causal/rational integration of the personal beliefs. I’m not saying that the personal belief set must be coherent, only that there are mechanisms in place that try to reduce incoherence. We are imperfect and sometimes irrational creatures. My point is merely that beliefs within the personal sphere resist conflict with other beliefs in that sphere in a way that they don’t resist conflict with subpersonal beliefs. What’s supposed to be impressive here is not the integration of the personal, but the isolation of the subpersonal. Thus, I suggest, to phrase things loosely, that the person-level beliefs just are the members of a set of beliefs closed under the “doesn’t like to conflict” relation. I choose a name for that relation that’s a bit unserious, as a reminder that it’s largely a placeholder until we fully understand the various ways in which person-level beliefs resist conflicting with each other. But the idea is: start with any conscious, occurrent belief; add to it any belief whose co-presence triggers the aforementioned compensatory mechanisms; reiterate with these just added beliefs and continue until there are no more beliefs to add: the resulting set comprises the person-level beliefs of the agent. This set might contain members that aren’t *directly* integrated in the relevant sense—that don’t, without the aid of other members of the set, stand in the “doesn’t like to conflict” relation to each other (maybe my beliefs about mental content conflict with my beliefs about skepticism, but I haven’t noticed the connection). But for any two members, I suggest, there will always be some path from one to the other, by way of intermediate beliefs that *are* directly integrated.

Consider some difficult cases: First, I might have such an emotional investment in believing that *p* that no contrary evidence would shake me from it. Yet, at some level I recognize that it is irrational, so I try not to use it as a premise in reasoning. Nonetheless it’s *my* belief; it is epistemically evaluable and its unjustified status would affect the justificatory status of beliefs that causally depend on it. Yet it’s in some sense causally isolated. However, the isolation here is different from the kind involved in modularity, since there’s a quarantine being actively sustained by the agent, rather than forced by the cognitive architecture. This wouldn’t be a counterexample to my proposal; it is an instance of the kind of compensatory mechanisms mentioned earlier. Note that the fact that this belief is nonlabile is no objection to its being personal. No subpersonal beliefs are labile, but not all personal beliefs are labile.

Second, the literature on implicit bias suggests that most of us have internalized a number of unjustified and highly racist (/sexist/etc.) assumptions about various groups. These assumptions are introspectively inaccessible to us, yet they are *our*, person-level beliefs. When they serve as premises for some other beliefs, their being unjustified detracts from the justification of the beliefs based on them. Suppose that after reading through some job applications, Candidate A seems unqualified to me, although I can’t point out which features of her resume are affecting my sense of qualification. A Reidian response to this, in the spirit of phenomenal conservatism, might be to say that my belief that Candidate A is unqualified is prima facie justified whenever based on this seeming—*no matter why it seems to me thus*. The view I’ve been defending here holds that whether I’m prima facie justified in this case depends in part on whether my belief in A’s qualifications depends on my (*my!*) unconscious and prejudicial beliefs about A’s ethnic group. Is my claim that these racist beliefs are personal compatible with what I’ve been claiming about isolation and integration?

I think so. In some ways, the assumptions responsible for implicit bias resemble the intramodular assumptions. They can conflict with the overt, personal beliefs of the agent (when they don’t, it’s not *implicit* bias), and they often, I presume, need to be inculcated by a long and repetitive process of indoctrination. But there are differences. First, the training of modules needs to be data-driven, while the indoctrination involved in bias doesn’t. Simply hearing my drunken, racist uncle spout nonsense about the location of light sources—even thousands of times—won’t affect my visual system’s assessments of convexity. But if his nonsense is about ethnic groups, it can have an effect on me, despite my better judgment. Second, the reasons why repetition is needed are different. It is for *architectural* reasons, to be discussed presently, that intramodular assumptions aren’t modifiable in response to a single instance of testimony. The reason the biasing assumption needs repetition to become established, however, is precisely that it conflicts with the overt, personal beliefs of the agent. Because of the conflict-minimization mechanisms already discussed, this makes the biasing assumption harder to establish. It can’t get in the front door, so it has to gradually chip its way in through the back. Again, it’s not a counterexample to my claims but an illustration of the compensatory conflict-reduction mechanisms.

Third, I might have certain logical incompatibilities in my belief system that I have long ago resigned myself to just live with. Sometimes this will involve the conflict-minimization mechanisms being in a low state of alert, on the lookout for a solution. Other times, I might be quite comfortable with it. I might think, for example, that it’s perfectly rational for me to believe that some of my beliefs are false; and I might be aware of and completely unbothered by the fact that this entails inconsistencies.[[13]](#footnote-14) I would think that this comfortable resignation only comes after an initial period of strain and dissonance, and that the comfort is only made possible by some additional beliefs, like beliefs to the effect that *certain* logical inconsistencies---like the ones involved in preface paradox-type cases---are epistemically innocuous. Again, this is a matter of the compensatory mechanisms doing their job and putting up an interposing belief to render the conflict tolerable.

Opacity, encapsulation, and the distinctive mode of learning are logically distinct, yet they tend to co-occur. Is there some reason for this co-occurrence? Here’s a hypothesis: we know from studies of connectionist networks that information can be implicitly encoded in a system without being explicitly represented. In a typical connectionist network, transient, occurrent representations are constituted by distributed patterns of activity of the simple processing units that make up the input, output and (if present) hidden layers. How the network processes the information—what inputs are mapped to what outputs—is determined by the pattern of weighted connections among the units. The weight matrix (the pattern of excitatory and inhibitory connections and their degrees of connectivity) thus embodies the system’s relatively stable knowledge of or assumptions about the domain. A connectionist implementation of a visual system, for example, would encode the assumptions we’ve been discussing (e.g., that things are lit from above, that retinally adjacent points are probably roughly equidistant, etc.) in the weights, rather than in the changing activation values of the units. The information encoded in the weights is fully distributed, in the sense that for any given piece of information (e.g., that the light source is overhead), there’s no proper subset of connections that encodes it; and for any weight or set of weights, there’s no principled allocation of some percentage of that weight to one assumption and some percentage of it to another assumption. Nor is there discreteness of causal efficacy: the whole body of knowledge is causally responsible for every processing event.[[14]](#footnote-15) I say this is implicit encoding rather than explicit representation because of the holistic way the information is realized in the weight matrix. The distributed representations of the layers have a rudimentary combinatorial syntax and semantics (the content of a pattern, or “vector,” is a function of the contents of the units that compose it), while the weight matrix clearly does not.[[15]](#footnote-16)

Because of the distributed way this information is encoded, programming these networks is a delicate affair. Programming is effected by training the network: running it on a large number of examples and slowly adjusting the weights appropriately. A network that is trained to, say, recognize cats cannot normally be then trained to recognize mice without thereby losing the ability to recognize cats, unless the network continues to train on cats. This is because the mouse training would require tweaking virtually all of the weights, writing over the weights that had encoded the cat information. The only way to avoid this “catastrophic forgetting” is to set up the network so that large changes to the weights are impossible, and to train the network on new tasks quite slowly, gradually, and through vast repetition, while continuing throughout to train it on the old tasks. This makes a distributed system incapable of fast learning—you can’t just tell it *p* and have it now know that *p* (if *p* is background information about the domain, rather than transient information about the current input)—but it also makes the system resistant to forgetting and to being misled by new information or misinformation.

Because the long-term information in the system is implicit in the weight matrix, rather than explicitly represented, it is difficult for downstream processes (or cognitive scientists, for that matter) to know just what information is encoded; one can’t just look and see what representations are activated, but must infer it from patterns of input-output mappings. This explains why a distributed network would be introspectively opaque, at least with respect to the assumptions implicitly coded in its weights. Furthermore, a system of this sort would not quickly modify its processing in response to relevant information about the domain: even if I learn that things here are lit from below, this won’t affect visual processing, in part because I can’t change the weights in my visual system in response to this new knowledge. In short, a distributed network will be at least partially encapsulated. It will also be, diachronically, a system that can only learn in a distinctive, gradual, way. The result is that information implicitly encoded in the weights of a distributed connectionist network will be insulated from, not integrated with, the person-level assumptions. Because of opacity and encapsulation, there will be little or no tendency toward the deployment of any of the compensatory mechanisms that I have suggested are indicative of the ‘doesn’t like to conflict’ relation. So, any system that is directly realized in a distributed connectionist network of the sort I’ve been describing will exhibit the central features of (weak) modularity. The assumptions implicit in its connection weights will also be subpersonal in virtue of these very same features.

**5. Cartesian Beliefs**

I am claiming that for a belief/assumption to serve as evidence, it must be a person-level belief, and it must be causally implicated in the agent’s having the justificandum belief. The kinds of beliefs that a Cartesian epistemologist invokes are plausibly person-level beliefs, but it is far from obvious that they are causally implicated, especially in light of the modular origins of many justificandum beliefs. I do normally believe that perception is reliable, and that I’m having such-and-such a visual experience, and that the thing in front of me has certain looks-properties; but it’s not clear that my ordinary perceptual beliefs counterfactually depend on my having these “Cartesian beliefs,” much less that they causally depend on it. Given the modularity of perception, any causal role played by person-level beliefs would have to be a sustaining, rather than initiating one; all the initiating doxastic causes are subpersonal. The idea would have to be that although we could *come to* our ordinary perceptual beliefs without the Cartesian beliefs, we wouldn’t *retain* them without the Cartesian beliefs. This weakens the Cartesian view significantly.

I stated in Section 1 that there was no in-principle objection to the claim that the whole belief set is causally implicated in the formation of every new belief. Our recent discussion of holistic implicit coding in connectionist networks explains how this might be true. But I warned that it is an empirical question which beliefs are in fact causally implicated. The beliefs that result from modular processes are not influenced by person-level beliefs, at least not in their formation. The Cartesian view could stay in business by dropping any claim about formation and only claiming that Cartesian beliefs are causally implicated in the retention of the relevant beliefs. But the view loses much of its plausibility, since surely one attraction to the Cartesian view was that some Cartesian beliefs seemed necessary to explain how perception moves past subjective appearances to beliefs about mind-independent objects. A worse problem is that it’s hard to see what would count as evidence for the Cartesian view, thus revised. It’s not obvious that many beliefs resulting from modular processes depend counterfactually on some Cartesian beliefs—our hunches about what would happen were we to believe that perception was *un*reliable, for example, don’t tell us much about what would happen were we simply not to believe that perception is reliable. But even if the counterfactual dependence is pervasive, we saw in Section 3 that that’s little or no reason to think these beliefs *casually* depend on Cartesian beliefs.

Worse yet, it’s not even clear how to conceive of the kind of causal dependence required. It’s easy enough to understand how implicit content could be causally implicated when that content is part of or embedded in a belief-producing process or mechanism. And it’s easy to understand how we might make sense of a distinction between initiating and sustaining causes in terms of belief-producing mechanisms: audition led me to believe there was a car coming from my left; when I turn and look, I see the car. Vision sustains this belief in the sense that vision is now outputting a belief that I already had. Had audition not been working, vision would have initiated this same belief. In the case of the beliefs that are produced by modular processes, it is rarely or never plausible that some extramodular process or mechanism would have sufficed to initiate these same beliefs, had the relevant modules not been operative. This doesn’t mean that the avowed counterfactual dependence *couldn’t* also be a causal dependence; it’s just that we lack a good model for understanding the causal mechanism.

In the end, however, the role of Cartesian beliefs will turn out to be an empirical matter. I think this is as it should be, although vested Reidians and vested Cartesians aren’t likely to be happy about it, for it makes these views hostage to an empirical hunch about causal relations, rather than something that can be fairly well established on the basis of what we already know. It might even turn out that the roughly Cartesian view is true for some cases but not for others. If, say, perceptual belief is modularly produced, but testimonial belief is post-modular, then there would be room in the testimonial belief-fixation pathway for person-level beliefs to exert a causal influence, even in the initiating of the testimonial belief. Testimony beliefs might then turn out to depend on Cartesian beliefs, while perceptual beliefs did not.

Even for the beliefs that we intuitively take to be perceptual, it’s an open, partly empirical question which ones are epistemically inferential and which are epistemically basic. For example, there are people in my town who I see regularly enough to recognize their faces, although I don’t know their names. Something like this is presumably going on: when I see a familiar person, I form the belief “there’s S2205,” where ‘S2205’ is a term in Mentalese that functions as a proper name, to pick out this individual. Surely I have such beliefs; they can be true or false, justified or unjustified, etc. But this belief doesn’t translate into English, because the only English approximation I could give is in terms of some definite description, and I’m not thinking of the person under any such description. Now I learn that this person’s name is Melissa, and the next time I see her, I think, “there’s Melissa.” This belief has all the phenomenal hallmarks of an epistemically basic belief: it’s immediate, compelling, I have an associated experiential state, I’m not aware of having formed any other beliefs that might serve as premises. But is it noninferential/basic?

The English belief ascription ‘there’s Melissa’ is ambiguous between two importantly different beliefs: one that is false if I have mistaken an unfamiliar lookalike for someone I’ve seen several times before, and one that is false if I was wrong about that familiar person’s name being Melissa. To disambiguate, we might call these two beliefs the belief that S2205 is here, and the belief that this person is named Melissa, respectively. This isn’t how I experience either belief, of course. ‘S2205’ is a conventionally assigned metalinguistic tag for a Mentalese expression and doesn’t capture the intrinsic, introspectible character of the belief. And when I make these sorts of judgments, I don’t find myself thinking about names, or expressions, or what people or things are called, even when the belief is one that I would readily retract were I to learn that I was using the wrong name. Nevertheless, there are two different beliefs I’m forming here, the belief that S2205 is here and the belief that someone named Melissa is here. I think that the former is a genuinely perceptual belief and the latter is not. The belief that the person here is named Melissa is based on my (person-level) beliefs that this is S2205 and that S2205 is named Melissa. This is seen from the fact that the belief that this person is named Melissa would be unjustified if either the belief that this is S2205, or the belief that S2205 is named Melissa, were themselves unjustified. Thus, on at least one disambiguation, the belief that Melissa is here is epistemically inferential.

What about on the other disambiguation? It depends on whether the belief that this is S2205 is based on any person-level beliefs. This is an extremely difficult issue, partly empirical and largely theoretical, concerning the contents of the outputs of the modular processes.[[16]](#footnote-17) Maybe the highest level attributive tokened by the perceptual system is one that represents the face as belonging to a particular person—this is the ‘there’s S2205’ representation I’ve been discussing; it would be true if Melissa is present but false if it’s her twin sister I’m seeing. But maybe it only represents the face as a face, having certain features; this one would be true even if it were her twin sister. Then any judgments about particular individuals would be post-modular and thus presumably nonbasic, evidentially dependent on the beliefs coming out of the modules. It might even be that the basic belief here represents the face merely as a physical object with certain perceptible properties; this would be true even if I’m looking at a mask. These are difficult issues that I can’t begin to answer here. To some extent, they’re empirical issues about how far encapsulation and introspective opacity extend, but they’re also about the nature of mental representation and the conditions under which a state has the content it has.

The status of Cartesian beliefs as evidence is, I think, up in the air. There is not currently good reason to think that Cartesian beliefs contribute evidentially to doxastically justified belief. But the ultimate case for or against the evidential role of Cartesian beliefs will depend on what kind of Cartesian belief we’re thinking of. My three main examples have been beliefs about source reliability, about what kinds of mental states the agent is in, and about what kinds of low-level or relational features some distal object has.

BonJour (e.g., 1976, 2010) has always held that Cartesian beliefs of the first sort play an indispensable role in the justification of psychologically immediate beliefs. We don’t know whether our beliefs that result from testimony are directly produced by modular processes; if not, then it’s likely that some kind of Cartesian beliefs are part of our evidence for the things we believe on the basis of testimony. But since perceptual and other beliefs surely are modular in origin, the proponent of this kind of Cartesianism has some work to do to make sense of a causal—as distinct from merely counterfactual—dependence that makes the *retention* of these beliefs causally dependent on person-level beliefs about reliability, even though the *production* of these beliefs is admitted not so dependent. Insofar as perceptual modules produce representations of objects and their properties, similar considerations hold for the second kind of Cartesian belief: personal beliefs about my own experiences aren’t needed to produce perceptual beliefs, and if they’re needed to retain them, we’d want to see an argument for that.

The third kind of Cartesian belief might be involved in the production of some psychologically immediate beliefs, even if these beliefs are modular in origin. One theory about the contents of perception (Glüer 2009) argues that perceptual experiences have a “phenomenal semantics”: rather than representing things as being red, they represent them as *looking* red, in the phenomenal sense of ‘looks’ (in contrast to a comparative or epistemic sense of ‘looks’, i.e., it describes the character of the experience the thing produces). Looking red is a real property of objects, but it’s a relational—indeed a phenomenal—property, one that holds in virtue of the phenomenal character of the sensations they produce in us. If the outputs of visual modules have a phenomenal semantics, then it is presumably this person-level belief about looks, along with one’s person-level belief that things that look red generally are, that justify one’s belief that the thing is red. McGrath (forthcoming) defends a similar view, although the looks properties he appeals to are not phenomenal properties of objects but, rather, the kinds of low level properties mentioned a few paragraphs back in connection with the face recognition example.

I’m not endorsing any kind of Cartesian theory; in fact, I suspect they’re all false, at least regarding IPM and System 1 beliefs. But I’m not arguing against them here. To argue *for or against* them would require a subtle and detailed argument, largely empirical in nature, but also thick with philosophical details about what’s involved in causally sustaining a belief or about how to assign contents to mental representations.

**6. Virtues of This Proposal**

I have argued that a belief/assumption can be part of someone’s justifying evidence for some belief only if that belief/assumption is both causally implicated in the holding of the target belief and is genuinely a belief *of the agent*, i.e., a person-level belief. I want to close by summarizing and elaborating what I take to be the virtues of my proposal.

First, it seems to get the cases right, not just the sunset and Jeep case that we started off with, but the resume evaluation case and the face recognition case as well. Where personal but unconscious beliefs are unjustified, the Reidian view would have us ascribe justification to beliefs that we ought to count as unjustified.

Second, it respects a broadly internalist constraint on evidence, without committing to a more general internalist epistemology. That is, it restricts one’s evidence to mental states *of the agent*—in a fairly demanding sense of the latter phrase, although it is neutral about an overarching evidentialism or the question whether external considerations like reliability might determine what’s evidence of what.

Additionally, the proposal fits well with the plausible idea that justification is determined, in part, by how well my belief fits with *my* evidence—not with *your* evidence, or *the* evidence in some detached, disembodied sense. That there is evidence for *p* doesn’t justify my believing that *p*, unless it’s evidence that I possess, and possession seems best understood in terms of mental states of mine.

Fourth, it fits well with a view I’ve defended elsewhere (Lyons 2009) about the relationship between modularity and basic beliefs: roughly, that a belief is basic just in case it’s the output of a modular system.

Finally, the proposal developed here allows us to make nuanced judgments about otherwise puzzling cases, even if those judgments are tentative and conditional. It doesn’t simplify our verdicts—the Reidian criterion would do that—and in fact, it suspends these verdicts until certain empirical/theoretical facts are determined, as we saw in the case of face recognition. What we saw there generalizes to other kinds of perceptual beliefs, and the solutions to the problems posed are not going to come easy. But this isn’t a defect of my proposal; it’s a virtue of the proposal that it allows us to see that the epistemology of perceptual belief always was more complicated than it might have appeared.[[17]](#footnote-18)

**References:**

Block, N. (2014). “Seeing‐As in the Light of Vision Science”. *Philosophy and Phenomenological Research* 89, 560-572.

BonJour, L. (1976). “The Coherence Theory of Empirical Knowledge”. *Philosophical Studies* 30, 281-312.

—. (2010). *Epistemology: Classic Problems and Contemporary Responses,* 2nd ed., Lanham, MD: Rowman and Littlefield

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

—. (2014). “Reply to Block: Adaptation and the Upper Border of Perception”. *Philosophy and Phenomenological Research* 89 (3): 573-583.

Carroll, L. (1895). “What the Tortoise Said to Achilles”. *Mind* 4, 278-80.

Conee, E. & R. Feldman (2004). *Evidentialism*. Oxford: Oxford University Press.  
Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, Mass.: MIT Press.  
Foley, R. (1987). *The Theory of Epistemic Rationality*. Cambridge: Harvard University Press.

Glüer, K. (2009). ”In Defence of a Doxastic Account of Experience." Mind & Language 24.3: 297-327.

Heck, R. (2000). “Nonconceptual Content and the Space of Reasons”. *Philosophical Review* 109, 483–523.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Lyons, J. (2001). “Carving the Mind at its (Not Necessarily Modular) Joints”. *British Journal for*

*the Philosophy of Science* 52: 277-302.  
—. (2009). *Perception and Basic Beliefs*. New York: Oxford University Press.  
—. (2014). “The Epistemological Import of Morphological Content”. *Philosophical Studies* 169 (3): 537-547.

—. (2015). “Unencapsulated Modules and Perceptual Judgment”. In J. Zeimbekis & A. Raftopoulos (eds.), *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford University Press (103-122).

—. (2016). “Inferentialism and Cognitive Penetration of Perception”. *Episteme* 13: 1-28.

—. (forthcoming). “Experiential Evidence?” *Philosophical Studies*.

McGrath, M. (Forthcoming). “Looks and Perceptual Justification”. *Philosophy and Phenomenological Research*.

Ramsey, W., S. Stich, & J. Garon (1990). “Connectionism, Eliminativism, and the Future of Folk Psychology”. In J. E. Tomberlin, ed., *Philosophical Perspectives* 4, 499-533. Atascadero, CA: Ridgeview.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group (1986). *Parallel Distributed Processing, vol. 1: Foundations*. Cambridge, MA: MIT Press.

Sellars, W. (1979). “More on Givenness and Explanatory Coherence,” in *Justification and Knowledge*, George S. Pappas, (ed.), Dordrecht: Reidel.

Senor, T. (2009). “Epistemological Problems of Memory”. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/memory-episprob/

Stanovich, K. and R. West (2000). “Individual Differences in Reasoning: Implications for the Rationality Debate”. *Behavioral and Brain Sciences* 39, 645–726.

Wu, W. (2014). *Attention*. New York: Routledge.

1. The Reidian Criterion as here formulated only prohibits unconscious beliefs from playing an *indispensable* evidential role; that is, even if they confer evidence on the psychologically immediate beliefs, these latter would have been justified without them. But proponents of the criterion typically deny any evidential role to unconscious beliefs. My arguments here will only concern the indispensable role, not any overdetermining one. [↑](#footnote-ref-2)
2. For more details, see Lyons (2009, forthcoming). [↑](#footnote-ref-3)
3. Maybe it’s not just any old belief in validity that does it; maybe it has to be an especially deep-seated belief, or an innate belief, etc. [↑](#footnote-ref-4)
4. For the record, I deny that experiences can serve as justifying evidence for beliefs (Lyons 2009), but I won’t be making any assumptions about that here. [↑](#footnote-ref-5)
5. It could still play an *epistemic* role without playing a specifically *evidential* one. [↑](#footnote-ref-6)
6. Recall the terminological note at the beginning of the paper. Although I think that ‘assumption’ is a better term here than ‘belief’, I’m not denying that these assumptions are a kind of belief. I am about to argue that they don’t play the same evidential role as (ordinary) beliefs, but that’s not going to be argued on the ground that it seems unnatural to *call* them beliefs. [↑](#footnote-ref-7)
7. This is a point I have belabored in a number of places, most notably Lyons 2009, 2014, and 2016. [↑](#footnote-ref-8)
8. One objection would claim that the assumptions embedded in subpersonal modules are not strictly speaking beliefs, and thus aren’t subject to EJB, and that they have the same evidential status as experiences. These assumptions, on this view, would be (like experiences) neither justified nor unjustified but still able to confer evidence (like experiences). This response would not adequately address epistemic defeat. Suppose I justifiedly believe that I’m not having an experience as of red, or that I’m in a situation where an experience as of red is no indication that something nearby is actually red. If my experience as of red is part of my evidence for believing that there’s something red nearby, then either of these just mentioned beliefs would undermine that experience’s ability to confer justification. Those justified beliefs of mine that conflict with the assumptions buried in my perceptual modules, however, do not generally undermine any justification. So the analogy between assumptions and experiences breaks down. If the epistemic contribution of the assumptions is not defeasible in light of justified beliefs, then that contribution is not an evidential one. [↑](#footnote-ref-9)
9. This is sometimes (Wu 2014, Cf. Burge’s 2010 discussion of primitive agency) articulated in terms of proper parts; I don’t necessarily reject this formulation, but mine is intended to be somewhat less committal. Although I contain my parts, I might also contain things that aren’t exactly parts of me—including, perhaps, some processing modules. [↑](#footnote-ref-10)
10. For the version of modularity I have in mind, which I take to be weaker and more plausible than the standard Fodorian one, see Lyons 2001, 2015, 2016). [↑](#footnote-ref-11)
11. Presumably *proximal* stimuli would do: a brain in a vat could engage in perceptual learning without any distal stimuli. [↑](#footnote-ref-12)
12. For example, learning to discriminate degrees of line slant in a certain quadrant of the visual field improves performance only for line slants, only for degrees within the range of training, and only in that quadrant of the visual field. [↑](#footnote-ref-13)
13. Thanks to Ernest Sosa for this example. [↑](#footnote-ref-14)
14. For an introduction to connectionist networks, see Rumelhart and McClelland (1986); for a discussion of the distributed nature of the information in the weight matrix, see Ramsey, Stich, and Garon (1990); for a discussion of implicit representation in connectionist networks aimed at the present concerns, see Lyons (2014). [↑](#footnote-ref-15)
15. I mean to be making the uncontroversial point about the great difference between the way that connections encode information and the way that pools of units do. I don’t mean to be taking any serious stand, here, on what counts as explicit or as genuinely representational. [↑](#footnote-ref-16)
16. See Block 2014 and Burge 2014 for an important exchange on this topic, though they are focused on where perception ends rather than on where modularity ends. I’m inclined to think (Lyons 2009) these are the same. [↑](#footnote-ref-17)
17. Earlier versions of this paper were presented at the Bled Philosophical Conference and the Southeastern Epistemology Conference. Thanks to audiences there, especially Brendan Dill and Terry Horgan for helpful comments. [↑](#footnote-ref-18)