Algorithm and Parameters: Solving the Generality Problem for Reliabilism

Jack C. Lyons

University of Arkansas

Forthcoming in *Philosophical Review*

ABSTRACT: I offer a solution to the generality problem for a reliabilist epistemology, by developing an "algorithm and parameters" scheme for type-individuating cognitive processes. Algorithms are detailed procedures for mapping inputs to outputs. Parameters are psychological variables that systematically affect processing. The relevant process type for a given token is given by the complete algorithmic characterization of the token, along with the values of all the causally relevant parameters. The typing that results is far removed from the typings of folk psychology, and from much of the epistemology literature. But it is principled, empirically grounded, and shows good prospects for yielding the desired epistemological verdicts. I articulate and elaborate the theory, drawing out some of its consequences. Toward the end, I apply the fleshed-out theory to two important case studies: hallucination and cognitive penetration of perception.

KEYWORDS: reliabilism, generality problem, justification, cognitive processes, naturalized epistemology, cognitive penetration

Probably the most significant development in epistemology in the last century or more is the development of a reliability theory of justification. The flagship version of the view is process reliabilism, which holds that a belief is prima facie justified just in case the cognitive process that produced it is a reliable one. But what counts as the cognitive process? The famous "generality problem" contends that there is no principled level of generality at which to type-individuate the cognitive processes responsible for a belief, without which

reliabilism fails to deliver any epistemic verdicts about any particular, even hypothetical, beliefs.[1] My current belief that there's a coffee cup in front of me is a token of many process types: perception, vision, perception in a noisy environment, true belief formed on a Thursday, cognition, and many others.[2] Clearly these have different degrees of reliability, so unless the reliabilist can specify a unique "*relevant*" process type (the type whose reliability determines the degree of justification for the belief) for each belief, it is unclear whether reliabilism says *anything at all* about which beliefs are justified and how much (Feldman 1985, Conee and Feldman 1998).[3] One worries that the reliabilist is applying the theory to particular cases by consulting her intuitions of justifiedness, cherry-picking a process type with a matching degree of reliability, and citing that type to explain the justificatory status of the belief. Such an explanation is obviously worthless. For these reasons, the generality problem has, for nearly forty years, been widely and rightly viewed as one of the most serious objections to process reliabilism.

I think that there is a very natural way for reliabilism to type-individuate cognitive processes, a way that is principled and which promises to give intuitively acceptable answers about which beliefs are justified. The individuation scheme we get can then be used to solve outstanding epistemological problems that have nothing to do with the generality problem. The theory I will propose is an "algorithm and parameters" (A&P) scheme for process individuation. The algorithm-individuation scheme is taken whole cloth from an already existing taxonomy in empirical cognitive psychology, but we will need to add parameters to the analysis to give us a typing that is narrow enough for our epistemological purposes. Parameters are also natural kinds recognized by empirical psychology, even though their role

[1] The generality problem was first articulated alongside the birth of process reliabilism itself in Goldman 1979 and was perhaps first pressed as an insoluble problem for reliabilism in Pollock 1984. It rose to prominence and received its name in Feldman 1985.
[2] For the sake of clarity, I'll use underlining to mark off candidate process types.
[3] Although reliabilism is often formulated in simpler terms—as I've done myself above—I will take it to be committed to the claim that the *degree* of justification is determined by the *degree* of reliability of the process.

in process individuation means that the process typing developed for epistemology, though still picking out natural psychological kinds, will be finer than the typing commonly used in psychology.

I start with some preliminaries in section 1 and articulate the A&P proposal in section 2. Section 3 focuses on an initial assessment of the theory. In section 4, I explain how this typing differs from commonsense thinking about processes and from the kinds of typing sometimes invoked in the literature. In section 5, I show how A&P clarifies and highlights the role of processes in process reliabilism. In section 6, I add some optional theses to the core theory and use the result to assess two important case studies: hallucination and the cognitive penetration of perception; the applications here illustrate the power of A&P to solve epistemological problems other than the generality problem, while also demonstrating how rigidly constrained the theory is in its verdicts about the relevant process type. I end, in section 7, with some brief concluding comments.

1. Preliminaries

The problem I aim to solve should be sharply distinguished from two other problems. Reliability is determined by the combination of the relevant process and the relevant environment, so a full process reliabilist theory will need to type individuate *environments* in addition to *processes*. Suppose a particular process yields mostly true beliefs over here by these rocks, but not if I step away a few feet. Does it count as reliable here? What if I have to go a few miles before its reliability diminishes? To another planet? It's not obvious how to give a general answer to such questions.

One very significant obstacle to answering these questions, besides their intrinsic difficulty, is the fact that there is a great deal of disagreement even among reliabilists, regarding the type of environment *relative to which reliability is assessed*. On one view, it's

simply the environment in which the process is employed (Goldman 1979, Lyons 2013), but on other views, it's the actual-world counterpart of that environment (Goldman 1979), some two-dimensional composite of these two (Sosa 1993, Comesaña 2002), a "normal worlds" environment (Goldman 1986), the environment that fixes the contents of the states (Burge 2003), the environment for which the process was designed (Bergmann 2006, Graham 2012), or all the environments in a particular swath of possible worlds (Henderson and Horgan 2006), and so on.

In any case, a theory of process individuation is aimed at something quite different. The debates just mentioned are about what a reliability attribution amounts to; my question here is what we're attributing reliability *to*. Since the problem of individuating processes can be fruitfully addressed without wading into these internecine disputes about environments, it's a good idea to pursue these independently.

Secondly and similarly, I'll be developing what we might call a "horizontal" rather than "vertical" individuation of processes. It's a theory about which belief tokens are the outputs of a common process, not about how far back in time those processes go (Weatherson 2012). Is the relevant process involved in remembering that *p* a process that goes back to the initial formation and storage of the belief that *p*, or do we only care about the recent part? Do acquired processes need to have been acquired via a reliable process-acquisition process? Again, these are controversial matters distinct from the central issue here.

Someone could insist that 'the generality problem' does or should refer to the larger, overarching individuation problem, which includes the problems of environments and vertical process individuation. But I think of the generality problem as the problem of (horizontally) individuating cognitive process types, and I think this is the problem most often thought of as the generality problem, by reliabilists and opponents alike. For good reason. It is or will become clear (in sections 4–6) that this type individuation of cognitive

processes has important consequences for reliabilist epistemology quite independent of the generality problem. It will also become clear that many of the important epistemological issues that run up against the generality problem do so in virtue of insufficient clarity surrounding processes, even when the environment and vertical factors aren't at issue.[4]

2. An Algorithm and Parameters Theory of Process Typing

The reliabilist requires a delicate balance of specificity and generality. The guiding idea behind process reliabilism is that justification is determined by the *general* tendency of a process to yield truth. On the other hand, if degree of justification is determined by degree of reliability of process, we need a fine-grained typing, for we need at least as many process types as there are degrees of justification. The basic idea I want to develop is this: the justificatory status of a belief is determined by (a) the narrowest (i.e., most determinate) psychological process type that was causally operative in producing or sustaining that belief (the algorithm), along with (b) the values of the general psychological variables affecting processing (the parameters). Thus, the relevant process type is given by the the algorithm used in producing or sustaining that belief, and the parameters within which that algorithm is operating. Hence an algorithm and parameters (A&P) typing of cognitive processes. Both central concepts—algorithm and parameter—are in need of explication.

[4] Some other epistemologies lean heavily on an underdeveloped notion of processes. Nozick's (1981) theory of knowledge relies crucially on "methods," which seem relevantly similar to cognitive processes. Siegel (2013) wants to compare perceptual etiologies with inferential belief etiologies; this requires typing these etiologies, which I suspect are just processes under a different name. I would be happy if the typing I develop here were of use to other, non-reliabilist epistemologies that need a theory of methods, processes, or etiologies. But I have enough to do here without arguing that it fits the needs of these very different theories as well, and I certainly wouldn't think that it was a shortcoming of my theory if it only works well for reliabilism. (I'll surely be alienating even some reliabilists along the way, let alone others!) Similarly, it is sometimes claimed that the generality problem is a problem for all epistemologies, not just reliabilism. This claim means different things for different authors: Comesaña (2006) means it to be the claim that everyone needs a theory of basing; Bishop (2010) means that everyone needs to allow—and make sense of—inference of the form 'I formed the belief that p through a reliable process; therefore p'; others might mean that justification is a matter of following epistemic rules, but each act of believing falls under a number of different rules; etc. Again, I'd be happy if A&P were of any help to non-reliabilist epistemologists, but I won't handicap the current project by insisting that it be. We will see that getting clear on processes reveals process epistemologies to be strikingly different from traditional internalism, in ways that I think are right, but which are not likely to be embraced by the traditional internalist.

The algorithmic typing will be entirely farmed out to cognitive psychology. But this will be too coarse-grained for epistemological purposes, so I will distinguish these "*psychological* process" types (which we can read right off from the science's typing) from "*cognitive* process" types (which are further individuated by consideration of parameters). It's the latter, narrower, types that determine degree of justification. (Any mention of "processes" alone should be read as referring to *cognitive* process *types*, unless context indicates otherwise.)

2.1 Psychological Process Types

Cognitive psychology is already deeply concerned with the empirical question of which psychological process types there are. Exactly which types the science will deliver for us will depend on:

(i) the methodological and general substantive presumptions of the field (is it behaviorist, cognitivist, Gibsonian, etc.?),

(ii) which general factors the theory takes to be causally relevant (most humans are sensitive to ambient light but not magnetic field, or day of the week, etc.), and

(iii) the level of detail at which the theory finds differences to be less important than similarities.

The first two of these are clear enough for now (although I'll return to (i) in section 6 below), but (iii) needs elaboration.

Cognitive science is replete with empirically grounded claims and debates about the sameness and distinctness of processes (i.e., process types). Dead reckoning is assumed to be a different navigational process than piloting from landmarks (Collett and Graham 2004, Gallistel 2007, Müller and Wehner 2010); small numerosities (up to 3 or 4) are thought to be handled by a different process than that by which larger numerosities are (Feigenson et al. 2004, Dehaene 2011, Mou and vanMarle 2014); visual face recognition plausibly involves a

different process than visual recognition of non-face stimuli (Peterson and Rhodes 2003, Farah 2004), etc.

But this doesn't go on without limit. Although there might well be distinct visual recognition process for artifacts as opposed to natural objects (maybe with the former using a parts-based algorithm and the latter a holistic one (Biederman 1987, Strat & Fischler 1991, Gerlach 2009)), no one thinks—as an empirical fact—that the visual recognition of ukuleles involves a different process from the visual recognition of candles; similarly for the visual recognition of dogs vs. cats, or male vs. female faces. Assuming that the scientific debates are well grounded, this gives us a lower limit to the individuation of processes.

Claims in psychology about whether *x* and *y* are subserved by the same process or different processes are generally claims about whether *x* and *y* use the same or different algorithms, in a particular sense of this term (Pylyshyn 1984). Algorithms, in this sense, are step-by-step procedures for computing a function (i.e., mapping a set of inputs to outputs), where each of the steps is itself a function.[5] To take a familiar example, multiplication is a function, and there are different ways of computing it. One way to compute the product *m* x *n* is to simply consult a very large lookup table. Another way is to do what most computers do: set *x* to zero and then add *m* to *x*, *n* times. Another is to use the method most of us learned in grade school: with *m* and *n* written in Arabic numerals, we take the rightmost digit of *m*, multiply it (using a lookup table) by the rightmost digit of *n*, and write down the answer; then we repeat for the *next*-rightmost digit of *n*, writing down the answer below our last one, but now with all the digits shifted one space to the left, etc., and then adding the results. These three different "ways" of multiplying involve three different algorithms: a lookup algorithm, a successive addition algorithm, and a partial products algorithm, respectively. To give a

[5] Certain processes are sometimes called "algorithmic" to distinguish them from "heuristic" processes; the former are guaranteed to get the right answer, while the latter are not. This is *not* the sense of 'algorithm' intended here. There can be heuristic algorithms without threat of oxymoron, so long as we understand algorithms simply as the computational procedures for mapping inputs to outputs.

complete algorithmic characterization of some procedure, system, or computational event, we need to specify new functions (input/output mappings) *by means of which* the original function is effected, and doing the same with the new functions, and so forth, until we reach a level of analysis at which the functions are mapped by the hardware directly, without mediation by other functions.

When psychologists talk about difference and sameness among processes, it's not usually functions they're concerned with, but algorithms as here understood. There are a number of "dual process" or "dual route" theories that hold that we have more than one way of computing a single function. It's widely agreed, for instance, that we have two different "routes" to reading a word aloud: one route is direct lookup of the word in a stored lexicon, the other is to use grapheme-to-phoneme conversion rules (Forster and Chambers 1973, Coltheart 2005). For a number of tasks, we seem to have a fast and sloppy, "System 1" heuristic process, and a different, slow, conscious, and careful, "System 2" process for the same (Schneider & Shiffrin 1977, Evans and Frankish 2009, Kahneman 2011). They compute approximately the same function (i.e., the same input-output mappings), but they do so in different ways.[6] In general, it is these "ways" or "routes" that psychology is concerned to articulate, and it is these that psychology is distinguishing when it distinguishes one process from another. This is just talk about algorithms in different terms. Many of the high profile debates in psychology (witness connectionism vs. classicism (Fodor and Pylyshyn 1988, Smolensky 1988), propositionalism vs. pictorialism in the imagery debate (Tye 1991), etc.) are debates not about what input-output mappings are computed, but about the computational

[6] The functions won't be *exactly* the same, since, e.g., some inputs that one argument evaluation algorithm maps to VALID will be mapped to INVALID by the other. As is common in psychology, there is a bit of idealization here in characterizing these as the same function, even though they're only approximately the same. For the present purposes, all that matters is that *even when psychologists are treating them as the same function*, they insist on distinguishing the two processes on algorithmic grounds. I return to this briefly below, in section 4.

means by which they're computed. Thus, it is clear that psychology individuates psychological processes by algorithm:

> Process tokens **Γ** and **Δ** are ***of the same (narrowest) psychological process type*** iff the complete algorithmic characterization of **Γ** is the same as the complete algorithmic characterization of **Δ**.

2.2 Parameters

Psychological processes are too broad to do the work needed for epistemology. We want to be able to say, for example, that, everything else being equal, a visual belief formed in poor light would be less justified than that same belief would be if formed in broad daylight, even if it uses the same algorithm.[7] Reasoning while highly intoxicated might use the same algorithms as sober reasoning, but we still think there's an important reliability difference that needs to be captured. For epistemological purposes we'll need to individuate cognitive processes more narrowly than psychological processes.

I will say that a ***parameter*** is a general psychological variable that systematically affects processing, i.e., a variable the differences in value of which result in differences of processing, across a wide range of inputs. Parameters will therefore include such factors as attention allocation, number of distractors, vividness of (memory or perceptual) experience, strength of "feeling of knowing", similarity to stored category template, number of such stored templates, shape of tuning curve of detector, degree of contrast in retinal image, presence or absence of semantic priming, mood, arousal level, etc. There are many different ways in which these variables might "affect" or "result in differences" of processing. They might systematically influence semantic character, speed of processing, depth of processing,

[7] As a matter of empirical fact, vision in dim lighting conditions might trigger various signal boosting subroutines that aren't called in good light, or it might involve a different cue combination algorithm than in good light, etc. The point here is that *even if* it turns out to be the same algorithm in both conditions, we should allow a difference in cognitive processes.

time allocated to search, behavioral reaction times, match criteria, confidence of judgment, accuracy of judgment, etc.[8]

One might wonder how two processes that use the same algorithm could differ in their reliability, since they will still produce the same outputs in response to the same inputs. One important way is that a single algorithm might be highly reliable for a given *range* of inputs and unreliable for a different range. Parameters like lighting conditions presumably affect processing and hence reliability and justification in at least this way, the kinds of inputs that occur in dim lighting conditions differing in general ways from the kinds that occur in bright light, leading systematically to general differences in the kinds of outputs delivered, with corresponding differences in reliability. The parameters of most interest for reliabilism are the ones that affect reliability, but affecting reliability isn't part of what it is to be a parameter. For a successive addition algorithm for multiplication, the cardinality of the factors affects processing time (not so for a lookup table, and for a partial products algorithm, it's the number of digits in the base-10 representation of the factors that does this, rather than cardinality per se), so cardinality is a parameter, even though it doesn't directly or indirectly affect reliability.[9]

Parameters can also sometimes serve as additional causal factors beyond the inputs already taken into account in the algorithmic characterization of the process. Such parameters could affect processing by e.g., introducing or increasing noise or error: a given algorithm might yield more consistent outputs under conditions of sobriety and attentiveness than under conditions of drunkenness and distraction. This requires us to think of at least some functions

[8] Some of these parameters are "usual suspects" in epistemology. Arousal, drunkenness, and the like figure into what Sosa (e.g., 2015) calls the "shape" of a competence, and they're the sorts of factors that internalists are happy to invoke. Lighting conditions are part of Sosa's "situation" and play a role in Feldman's (1985) initial presentation of the generality problem. Others, however, like shapes of tuning curves, or degree of match to template, are unlikely to be mentioned except by process reliabilists (the latter is explicitly noted by Goldman 1986).

[9] This example shows that parameters will need to be relativized to algorithms. I return to this in section 2.3.

and algorithms as involving a stochastic element—otherwise a given input would have to return exactly the same output in every single case. I think this stochastic understanding of functions and algorithms has always already been at least implicit in psychology; conceiving real-life psychological transitions in algorithmic terms would be implausible otherwise.

Third, parameters might systematically affect processing by affecting which algorithm is used. Cognitive load, for example, can bias one toward the use of effortless System 1 algorithms instead of resource-intensive System 2 algorithms (Evans and Stanovich 2013), as can positive mood (Schwarz 2002). This is the least interesting role for parameters, however, since it is already covered by individuating psychological processes in accordance with algorithms.

There's an additional bit of work this conception of parameters can do. Psychological states are quite often the result of a number of distinct processes working together in a competitive or cooperative way. Visual depth perception, for example, takes several different binocular and monocular cues and calculates a weighted sum of the depth estimates from each of them (Howard 2002). That weighting (i.e., the specification of what weight is given to which cues in the summative process) is itself an additional psychological parameter, beyond lighting conditions, degree of retinal disparity, etc.

2.3 Avoiding trivialization

Differences of "mere content," as I will call them, do not ipso facto involve different processes, according to A&P. A dog-shaped proximal stimulus produces different psychological states than a cat-shaped proximal stimulus (*dog*, rather than *cat* representations). But this makes for neither a difference in algorithm nor a difference in parameters.

Algorithms are vulnerable to threats of trivialization, in the following way: functions can be represented as or equated with sets of ordered input-output pairs; therefore algorithms

can be represented as or equated with sets of *strings* of ordered input-output pairs. But any non-unit, nonempty such set can be divided into two or more subsets, yielding two or more distinct algorithms. Thus, the "process" that maps dog-shaped proximal stimuli onto *dog* categorizations will trivially involve a different algorithm from the one that maps cat-shaped proximal stimuli onto *cat* categorizations. More generally, *any* difference of either input or output will make for a difference in process.

If psychological processes were trivializable in this way, then all the debates about sameness or difference of process for different contents would be empty—they'd always be different processes, but trivially so. Since the debates aren't empty, such trivialization must be foreclosed somehow; algorithms can't be separated thus. But why not? I suggest that the debates presuppose that the individuation of algorithms is grounded in the "unitariness" of the realizing mechanism, even as these same debates are neutral about just what that the relevant mechanisms are. Following Lyons 2001, we can think of a computed function as a set of ordered input-output pairs and define a ***part*** of that function as any proper subset of that set. Function *F* is ***unitary*** with respect to some substrate, or mechanism, *S*, iff *S* is capable of computing *F*, and no proper part of *S* is capable of computing any part of *F*, in the absence of other computational mechanisms. The complete algorithmic characterization of a processing event must advert to functions and thus algorithms *that are unitary* with respect to whatever the mechanisms are. The idea, then, is that *if* you couldn't lesion the dog-recognition mechanisms without thereby also knocking out the cat-recognition mechanisms, then dog recognition and cat recognition don't involve distinct algorithms.[10]

[10] Psychological-level debates about algorithms are, of course, often carried out without much or any reference to or knowledge of the underlying substrates or counterfactuals about them. Nevertheless, it's an empirically very well-educated guess that you couldn't knock out dog recognition without knocking out cat recognition, or that you *could* knock out visual cat recognition without knocking out auditory cat recognition, and so forth. Thus, even though the psychological-level debates are in this way hostage to yet unknown lower level facts about mechanisms, this typically doesn't pose much of an epistemic threat.

Nor can a difference of mere content make for different parameters. The difference between cat-shaped and dog-shaped proximal stimuli, for example, isn't a *general* one; it doesn't concern a "wide range" of inputs. But now another problem arises, for this vague formulation suggests that parameters might be easily trivialized even if algorithms aren't. I think that we can sharpen the concept of parameters by appeal to the natural kind structure of psychological phenomena. I propose that parameters are only *those variables about which there are psychological laws*. We saw in section 2.2 that a variable (e.g., number of digits in the input) might have systematic effects relative to one algorithm but not relative to another. Thus, what we will really need is a relativized view of parameters:

> a psychological variable is a ***parameter*** relative to algorithm *A* iff there are psychological laws describing that variable's effects on processing, and at least one of those laws govern the operation of *A*.

The aforementioned examples of parameters (mood, degree of match to template, etc.) presumably figure into psychological laws, so all these variables would count as parameters, and any difference in their values will yield a difference in cognitive process type (relative to some algorithms, at least). There are presumably *not* laws covering the visual processing of cows, on a Tuesday, in Cambridge, or concerning true beliefs as such, etc. So these are *not* parameters, and differences in their values don't make for differences in cognitive process type.[11]

I won't argue here that this is the best way to save the psychological debates about algorithms from vacuity, except to note that cognitive neuroscience is likely independently committed to something like a unitariness constraint (Lyons 2001) and that even psychologists who don't care about physical realization want their theories to eventually integrate with cognitive neuroscience. All that is required for the present purposes is that it's principled and does indeed save algorithm individuation from triviality.

[11] Even causal factors won't make a difference to process type if they don't have a systematic (i.e., law-governed) effect on processing. An otherwise reliable process that's zapped by Q-radiation still counts as that same process, even if the Q-radiation is affecting it. Similarly, if it's not a matter of psychological law, but just a strange quirk, that one psychological state is affecting processing on a particular occasion, this won't make for a difference in process type. This, anyway, is what follows from the view I'm defending on other grounds. I don't have strong intuitions in favor of or against this consequence.

Any solution to the generality problem is going to face a looming threat of trivialization, although what form this threat takes will depend on the shape of the solution. Those, like A&P, that type processes very narrowly to capture fine-grained epistemic distinctions will have to show that the processes aren't trivially narrow. A&P claims that the lower bounds on algorithm and parameter individuation—and thus the solution to the trivialization threat—are already there in the psychology, though perhaps implicitly, in the form of unitariness and nomological constraints on parameters.[12]

2.4 Cognitive process types

So far, we have a view about algorithms (aka "psychological process" types) and a view about parameters. We get a theory of *cognitive* process types (these are the types directly relevant to justificatory status) by simply putting together what's been said about *psychological* process types with what's been said so far about parameters:

> Process tokens **Γ** and **Δ** are tokens of the same (relevant) cognitive process type iff the complete algorithmic characterization of **Γ** is the same as the complete algorithmic characterization of **Δ**, and **Γ** and **Δ** have the same parameter values.

Removing the comparative language, and keeping in mind that a belief can be the result of more than just one process, we get:

> The **cognitive process type** that determines the justificatory status of a belief token is given by the complete algorithmic characterization of every psychological process token causally relevant to that belief, along with the associated parameter values for all those processes.

[12] My solution to the trivialization worries in this section helps itself to the difficult concepts of counterfactuals (algorithms) and laws (parameters). It's not that I assume these concepts to be unproblematic. Rather, I assume they're non-epistemological problems, and problems that would need to be solved anyway, independent of A&P. For several obvious reasons, I'm not in a position to specify exactly what laws and hence what parameters there are; thus, I can't fully guarantee that we won't have too many or too few. Nevertheless, the clear and obvious instances of laws and of non-laws mentioned here and examined in more detail throughout the rest of this paper seem to capture pretty well those factors that we do and don't, respectively, want to incorporate into our epistemology.

Thus, for example, when I look at a cat, the cognitive process won't be seeing a cat, but something more along the lines of visual recognition for natural stimuli in lighting conditions *L*, with causally relevant retinal image size *S*, with attentional resources distributed in manner *A*, etc. (there are surely more parameters than these). When my uncle comes to believe that immigrants are planning to destroy America, the process won't be testimony, or watching TV, or watching Fox News, but something much more complicated. What that is exactly will depend on empirical facts that are currently unknown, as well as on details about the case that haven't been specified. There will be some perceptual process for hearing what is said, a language comprehension process for extracting the content (plus perhaps another one for decoding the implicatures), and some kind of testimony acceptance process, all operating under parameters specifying the testimony's degree of consonance with background beliefs, the subject's relevant motivational states, etc. The fact that it's Fox News figures neither into the algorithm nor directly into any of the parameters, but if my uncle feels a special affinity toward the channel, which influences his readiness to accept the testimony, that affinity serves as a parameter value. There may be laws to the effect that people grant more credulity to in-group members—or to white, attractive, blond women—than to other groups; that would give us another parameter. Maybe psychological processes can span more than one individual (Goldberg 2010), in which case some of the cognitive processes of the speakers might have to figure into our description of the full process.

Obviously, verbal descriptions of cognitive processes won't be nearly as pithy as the examples we usually read and talk about, and a great deal of empirical information will be needed even to know what the potential algorithms are and what kinds of variables make a systematic difference to processing in any given case.

Psychology is not normally very explicitly concerned with cognitive process types; it's much more concerned with psychological process types. The latter are presumably

natural kinds, but that doesn't mean that the former aren't also. Because of the role of psychological laws in constraining parameters and the role of parameters in specifying cognitive processes these will turn out to be natural kinds as well, even if they're not the natural kinds that are most prominently visible in empirical psychology.

2.5 Some related solutions to the generality problem

Although an exhaustive discussion of other solutions to the generality problem is beyond the scope of this paper, it will be clear that A&P—or the algorithm half of it, anyway—bears obvious and important similarities to process typings proposed by Goldman (1979, 1986), Alston (1995), and Beebe (2004), notably to Goldman's early suggestion that "the critical type is the narrowest type that is causally operative in producing the belief token in question" (1986, p. 50).[13] Goldman's comments on the problem are rather brief, so I'll focus on the other two. Readers familiar with these other views will see that the current proposal is quite differently motivated from and elaborated in more detail than these theories. Like Alston and Beebe, I'm developing a typing rooted in the sciences. Like Alston (and I believe Goldman and Beebe, though they're less explicit about it), I rely on psychological realism in some sense of the term.

Alston claims to individuate processes by function, rather than algorithm, although he doesn't think that any two input-output equivalent processes are necessarily of the same type, which suggests that he may have something more like algorithms in mind. At times, however, he seems to think that the relevant process-defining function for a given token is a mapping from just that particular input (or rather, the causally relevant aspects of that input token) to the particular belief output, in such a way that the function involved in recognizing a maple tree is not the same function that is involved in recognizing a cat, or perhaps even an oak tree.

[13] For some very different solutions to the generality problem—different enough that I won't address them here—see Heller 1995, Wunderlich 2003, Kappel 2006.

If so, his proposal might more closely resemble Comesaña's (2006) than mine. Comesaña holds that the relevant type for a belief that *p* based on evidence *e* is the type believing that *p* on the basis of *e*. I consider this a kind of indicator reliabilism, rather than process reliabilism, since the means by which the system moves from *e* to *p* is ignored, as is the more general reliability of the same mechanism regarding other contents (see sections 5 and 6.2 below for more on why these differences matter).[14]

Like Alston, I appeal to natural kinds in individuating processes, although he puts it to a different use and even means something quite different by it. He thinks that there is just one natural kind for each process token, something like a fundamental or essential kind (1995, 10-11). I am dubious of such an idea and certainly don't rely on it. I do claim that psychology cuts nature at (some of) its joints, but I don't assert or require that the kinds ascribed by psychology are any *more* natural than the kinds ascribed by, say, neuroscience or chemistry. I claim that my cognitive process typing reveals natural kinds but not that they're any more natural than those revealed by other typings.

Neither Alston nor Beebe say anything about how exactly to individuate algorithms or functions. They don't give us a way to tell whether the visual identification of dogs involves the same algorithm or function as the visual identification of cats, or whether the process of multiplying 7 x 8 could possibly be used to multiply other numbers. In particular, they don't recognize or address the trivialization worries of section 2.3. These worries are quite central, and without answers to them, the original generality problem very quickly reasserts itself, as Conee and Feldman (1998) note in their reply to Alston.

None of Goldman, Alston, or Beebe appeal to anything similar to parameters. Beebe's algorithms, like mine, require supplementation. But he thinks this for very different reasons

[14] The idea that Alston's 1995 paper may be more a defense of indicator reliabilism (which he defended in Alston 1988) than of process reliabilism would explain his otherwise quite puzzling insistence that the inputs to the process are experiences, rather than, say, sensory stimulations.

than I do, and he solves it with something quite unlike parameters. I invoked parameters because algorithmic typing doesn't give us all the epistemic distinctions we need. Beebe's concern is that his algorithmic typing will yield processes like perceptual process used by a middle-aged man on a Wednesday (this is not true of A&P's algorithmic typing, since cognizer's age and date aren't inputs). It is ostensibly to purge these irrelevant factors like age and date that Beebe supplements the algorithmic typing. What he supplements it with is "statistically relevant factors": two process tokens are of the same type, for Beebe, only if they use the same algorithm and are alike with respect to every factor that alters the probability of the process yielding a true belief.

Parameters differ significantly from statistically relevant factors. Not all parameters are statistically relevant (they might affect processing without affecting reliability), and not all statistically relevant factors are parameters (the fact that the output is true is statistically relevant, but it's not a causal factor that affects processing, so it's not a parameter). Truth and accuracy are important to psychologists, but they don't influence *processing*, i.e., the transitions from inputs to outputs. Statistically relevant factors are not among the criteria by which psychology or any cognate science type-individuates processes, so Beebe's processes will not turn out to be natural kinds, as they will on A&P (section 3, item 3). Also, distinguishing processes by statistically relevant factors will turn differences of mere content into differences of process, with epistemological results that I find unwelcome (section 5.1). In fact, as Dutant and Olsson (2012) show, Beebe's use of statistically relevant factors gives us a typing so narrow as to yield the extremely undesirable result that a belief is justified if and only if it's true.

Note that not every solution to the generality problem that looks different from A&P is a competitor to A&P. A&P is a theory about which processes determine the epistemic status of a belief. Some solutions to the generality problem (e.g., Jönsson 2013, Olsson 2016)

are aimed at something quite different: a theory about what drives ordinary *intuitive judgments* of justifiedness. These two—actual epistemic status and intuitive judgment about epistemic status—needn't have too much to do with each other, especially on an externalist theory that denies that first-order justification depends on a reflective metajustification. There are, of course, deep and complex metaepistemological issues concerning the connection between philosophical theories of justification and our intuitive judgments of justifiedness; I will briefly address some of this below in section 4.2.[15]

3. Assessing A&P

Fully fleshing out the A&P theory and applying it to particular cases will require a great deal of empirical information, much of which outstrips our current knowledge. But even from where we now stand, I think we can see that the theory has a number of attractive features.

1. *A&P specifies a unique relevant type*. It requires the finest algorithmic individuation possible, modulo unitariness of mechanism. It requires that processes be further distinguished by any difference in parameter values, where a parameter is any psychological variable that has a systematic (i.e., lawlike) effect on processing. As long as there are determinate facts about what algorithms are in use and what psychological laws there are, there will be determinate facts about which processes there are and about which unique relevant cognitive process type a given token belongs to. The unitariness constraint and the

[15] For what it's worth, I don't think that Olsson takes the trivialization worries of section 2.3 seriously enough. Olsson claims that "the single-purpose [i.e., relevant] category for token process *t* for taxonomy T (with respect to reliability) is the most inclusive T-category C such that (i) *t* belongs to C and (ii) C is associated with the same reliability attribute (reliable/unreliable) as is the most specific T-category to which *t* belongs" (2016, p. 190). If I understand it, the proposal is that, of all the types/categories to which *t* belongs, the one whose reliability determines justification is the broadest type that has the same reliability value as the narrowest one. But without substantive constraints on T-categories and their lower bounds (literally everything that Olsson says about T-categories in this paper is included in the above quotation, and I don't see constraints on categories more generally that might address the issue), the narrowest category will be the one that contains just *t*. This means that a belief will be be justified (or intuitively judged so) if and only if true.

nomological constraint on parameters ensure that the typing will bottom out at a certain level of fine grainedness. This point is in no way undermined by the fact that we might encounter or imagine cases the details of which are too sketchy for us to *know* what the relevant type is.

2. *The relevant type is determined by the theory,* not by the theorist, and certainly not with an eye toward getting the desired answer. A&P, as adumbrated so far, is somewhat general and schematic, leaving room for disagreement both empirical (can knowing that you're holding a tool in your hand influence visual estimates of depth (Witt et al. 2005)?) and theoretical (can distal states of affairs be among the inputs to genuinely cognitive processes, or do only proximal stimuli count?). Different theorists will thus have different views about exactly which processes are responsible for which beliefs. But these are very *general* disagreements and will normally require argumentation quite independent of the particular epistemological cases we're interested in at any given moment. Thus, there's very little room for the theorist to use A&P to generate post-hoc rationalizations for the epistemic verdicts she wanted on intuitive grounds.

3. *The typing is independently motivated*. The typing of *psychological* processes is farmed out entirely to psychology in an obvious way and is thus completely independent of any epistemological concerns. Our choice of psychological types, rather than, say, neurocomputational types is, of course, driven by epistemological interests, as is the role of parameters in the type individuation of cognitive processes. So our *selection* of this typing is driven by epistemological concerns, but the selection is of one among a number of already existent typings that carve the world at its non-epistemic joints. And even though psychology doesn't normally individuate processes as narrowly as A&P (since it normally types by algorithm only), cognitive processes as here individuated are still natural, psychological, kinds. It's psychological laws that guarantee the natural kind status of the typing, and they do so independently of epistemological concerns.

4. *The theory meshes with our best science*. Who wouldn't want that?

5. *The theory types processes at an appropriate level of generality*. A&P delivers a very fine-grained typing of processes. Any difference in parameter values makes for a difference in process, and some parameters may be continuous; there will thus be vastly many processes, surely enough for us to make the epistemic distinctions we need made. And yet each type will contain many possible tokens—tokens with other contents and other truth-values—so our processes have both the specificity and the generality we needed.

Some might worry that with continuous parameters many process tokens are going to be practically unrepeatable and thus the only actual instance of their type.[16] If for this or for other reasons, we don't want parameters to be continuous, we could develop various technical methods for quantizing these continuous variables, or of keying reliability to a range of values instead of a very precise value; these would give each type more actual instances. I won't pursue such methods here, although I will note that the solutions can be developed in general and in advance, so there's no threat that the details are being gerrymandered to get particular cases intuitively correct. Second, even if each process token were the only one of its type, this only threatens the distinction between reliability and truth on a dubious and unpopular understanding of reliability, according to which it is simply a matter of actual track record. We might just as plausibly claim that a newly minted coin, flipped once and then destroyed, must have been biased. There are serious controversies about how to understand the kind of environment in which reliability is determined (as seen in section 1), but all the standard views are compatible with taking reliability to be a matter of counterfactuals and not actual track record. One who insists on an actual frequency account of reliability, however, is free to pursue the technical methods just mentioned.

[16] Thanks to Julien Dutant and an editor for this journal for pressure on this point and discussion of it.

6. *A&P satisfies—at least approximately—our pretheoretic intuitions* about which beliefs are justified. I say "at least approximately" because even if our pretheoretic concept of justification is that of reliably formed belief, precisifying the concept of processes in this empirically grounded way is sure to yield a process typology that deviates from folk psychological categories. Thus, it would be surprising if there weren't *some* divergence between the verdicts of our theory and our folk intuitions. It is therefore more important that A&P delivers epistemic verdicts that are acceptable on reflection. Defense of A&P on this score requires discussion of many examples, which will concern much of the remainder of this paper.[17]

Consider a well-known example Conee and Feldman use to illustrate the generality problem:

> The token event sequence in our example of seeing the maple tree [through a window] is an instance of the following types, among others: [1:] visually initiated belief-forming process, [2:] process of a retinal image of such-and-such specific characteristics leading to a belief that there is a maple tree nearby, [3:] process of relying on a leaf shape to form a tree-classifying judgment, [4:] perceptual process of classifying by species a tree located behind a solid obstruction, etc. The number of types is unlimited (Conee and Feldman 1998, p. 2)

It's clear that none of these counts as a cognitive process, according to A&P. 1 is far too broad. 2 is both too broad, in that it doesn't specify which visual algorithms are involved, and too narrow, as it invokes mere content in both the input and output. Same with 3. 4 has the

[17] In what follows, I will focus on prima facie justification; this is what I take reliabilism primarily to be a theory of. If defeat cannot be understood in terms of process reliability but only in, say, responsibilist or evidentialist terms (Beddor 2015), then A&P can only be relevant to that prima facie component of justification. However, if defeat can be understood in terms of reliable processes (Goldman 1979, Lyons 2009, 2016b, Graham and Lyons forthcoming), then A&P will be relevant to the whole epistemological story. The details of a process theory of defeat, however, are complex and controversial, and they won't make any difference to the assessment of A&P, so I will generally ignore defeat.

same problems and also treats the window as if it were a psychological parameter, which it's not (nor is any real psychological parameter indirectly alluded to, unless the glass is stained, or frosted, or otherwise introduces some distortion that's reflected in the values of the psychological variables).

Similarly, Feldman (1985) asks about a particular example whether we should take the relevant process type to be "the perceptual process, the visual process, processes that occur on Wednesday, processes that lead to true beliefs, etc" (p. 159). Again, it is clear that none of these is a *cognitive* process type.

Instead, the relevant psychological process type in the maple tree case is some species of visual object recognition for natural items. If gist perception as of an outdoor scene factors into the computation, then that's part of the complete algorithmic characterization as well. Similarly if motion perception contributes to the discrimination and/or identification, etc. The parameters of the visual process include factors like image contrast, retinal size of the object, resolution of the relevant region (the part of the image that contains the tree), etc., as well as idiosyncratic features of the agent, like visual acuity, sharpness and sensitivity of visual matching templates, etc. If multiple independent algorithms are combined by weighted sum, the relative contribution of each is a parameter.

Insofar as these algorithmic and parameter factors affect reliability, it seems right—at least from a reliabilist standpoint—that they should figure into the determination of the relevant cognitive process type and hence the justificatory status of the belief.

More generally, we want for vision in good lighting conditions, up close, for familiar types of stimuli to involve different cognitive process types than vision in bad lighting conditions, from far away, for unfamiliar stimuli, even if the same algorithm is used, because we want to support an epistemic distinction between them. This is easy, as these two involve different parameters, and these parameters affect reliability in line with our intuitions about

justification. Careful, System 2, solution of logic problem while awake and sober should involve a different cognitive process type than from-the-hip System 1 solution of the same problem while drunk and chasing toddlers, since the degree of justification is clearly different in these two cases (a fact not obviously accommodated by other epistemologies). Again, this is easy, because these involve both different algorithms and different parameters. We want memory search for confirming evidence for a hypothesis to produce more justification if it's dispassionate than if it's heavily influenced by emotional investment in that hypothesis. We want a negative verdict in a case of visual search (e.g., "there are no numbers present in this array of letters") to have a degree of justification that's sensitive to the number of distractors, to the time it took to make the judgment, and to the attentional resources devoted to it. In all of these contrasting cases, A&P clearly specifies different processes, and the intuitive differences in the reliability of these processes track quite well the intuitive differences in the justification of the beliefs.

Note that we can make many such comparative claims even without knowing or specifying in full detail what each of the processes are. Although it's often hard to name the whole process responsible for any given belief (as seen in section 2.4), it's not always hard to name factors, in the form of parameters and of "sub-algorithms," (i.e., algorithmic subroutines) that should make a difference to reliability and thus to the justification of the belief.

4. Scientific and folk taxonomies

The processes we have arrived at are not the familiar entities that populate common sense, or even epistemology writings. This deserves comment.

4.1 Psychology and organon

Processes as understood by A&P are very different from the kinds of "methods" of belief-fixation we sometimes think of, especially in meliorative or social epistemological contexts. The difference I have in mind is between a ***psychology*** (a set of mechanisms of belief-formation) and an ***organon*** (a set of advisory rules for pursuing inquiry), respectively. Constructive dilemma, argument *ad hominem*, hasty generalization, and their ilk are organonic categories, but they're too abstract to be cognitive or psychological processes. The fact that theorists have often gone back and forth between the psychological and organonic vocabularies without apparent awareness of doing so has probably contributed a lot to the sense that reliabilists are individuating processes in an inconsistent and ad hoc manner.

The difference here is an important one. Some theories in the psychology of deductive inference invoke a kind of "inner logic," whereby we reason from premises in a formal, syntactic, rule-governed way, irrespective of the contents of the premises and conclusions (Braine & O'Brien 1998). Most psychologists, however, think we reason deductively by forming and transforming mental models (Johnson-Laird 2001), or by drawing analogical inferences from known cases (Shastri & Ajjanagadde 1993), or some other non-syntactic method. Evidence for the various theories consists mainly of the observation that we are better at some deductive problems than at others, which is measured primarily in terms of pervasive and predictable errors.

These non-syntactic theories, which explain our capacity for deductive reasoning in terms of heuristic processes, would be paradoxical if 'deduction' had to pick out a particular kind of psychological process, since deduction as logicians think of it is paradigmatically syntactic and paradigmatically non-heuristic. Instead, the word 'deduction' here must be ambiguous among problem (the thing to be solved), function (the actual input-output mapping), and algorithm (the means of effecting that mapping). The study of deductive reasoning in psychology normally uses the term 'deduction' in one of the former two senses,

to specify a *task-level* phenomenon (as I'll call it), rather than an algorithmic-level phenomenon, and the competing theories propose different algorithms for the task of deduction.[18]

The reliabilist thinks that the actual way by which the conclusion is reached matters quite a lot to the epistemic status of the belief. We may sometimes solve deductive problems by using a quick and dirty System 1 process whose inner workings are introspectively inaccessible and that trades some accuracy for speed by using heuristic processes. Other times we may use a slow but highly reliable System 2 process, whereby we deliberately and knowingly apply the rules of logic to the problem at hand. The epistemic status of a particular belief should depend on which of these two processes we are using on a given occasion, even though the inference was valid either way. In both cases, we are engaged in deductive reasoning in the task-level sense—indeed, drawing the same conclusion from the same premises—although only System 2 uses a deductive *algorithm*. The *task* is deduction in both cases, but the *belief-forming process* is deductive in nature (in the sense of being formal and non-heuristic) only in the System 2 case.

This is why 'constructive dilemma', 'argument *ad hominem'*, 'hasty generalization', and the like do not name cognitive processes. Like 'deduction', these pick out task-level phenomena, which might be executed by way of different algorithms in different individuals or on different occasions.

Although the organon is aimed at a higher level of abstraction than cognitive processes (namely, the task level), it might still play two important epistemic roles. First, it

[18] Problem and function are distinct because our actual confrontations with deductive problems don't always result in our computing deductive functions, but only approximations thereto, given the robust and recurring errors we make. Psychologists are often well served by not making too much of a fuss about the distinction between function and problem, however, since deviations between actual and ideal input/output mapping might be the result of noise, distraction, etc. (so-called "performance errors"), or they might be the result of using an algorithm that even at its best only computes a function that approximates the ideal ("competence errors"). It's best not to prejudge these issues going in. For this reason and for simplicity's sake I lump problem and function together using the appropriately ambiguous term 'task'.

gives us explicit rules that we might learn and apply. If I know that argument *ad hominem* is illicit, then this knowledge might serve as a defeater for me when I do it anyway. Learning rules for inductive generalization might give me premises that I can use in System 2 to form generalizations. Deductive and inductive inferences might then employ the same System 2 algorithm over their different contents, and yet differ in such parameters as number of premises, ease of inferential transition, felt certainty of conclusion, etc.

Second, although the organon doesn't *specify* belief-forming processes, it *sorts* them. It groups together broad classes of cognitive processes that share similar input-output mappings, and classifies them as epistemically good or bad. If arguing *ad hominem* is epistemically bad, then it's because a number of distinct process-level phenomena are therefore bad: downgrading or discounting someone's argument because of a personal dislike for the speaker is a process-level entity, or at least it's getting closer. More accurately, there is a certain standard suite of processes for evaluating argumentation; and emotional factors, like dislike for the speaker (or admiration, for that matter), can have systematic and predictable effects on these processes. When the emotion affects processing, it is a parameter whose influence makes for a distinct (and less reliable) process.[19]

The sorting, however, will be rough and approximate, as we will see soon (in section 5). A belief might satisfy (/violate) the organonic rules without being prima facie justified (/unjustified), even if most organon-conforming beliefs are the results of reliable processes and vice versa.

4.2 A&P versus intuitive individuation of process types

The typing A&P delivers is very fine-grained and esoteric. Ordinary folk, with no philosophical training and no erudite cognitive scientific knowledge, surely aren't forming

[19] Similar considerations apply to epistemic virtues, which also make an appearance in the organon. Openmindedness isn't a process, but the degree to which my assessment of an argument is influenced by a desire to be proven right is a parameter, one that presumably affects reliability and justification.

intuitive judgments of justifiedness on the basis of the kinds of processes described by A&P. Can a theory that incorporates A&P have anything to do with our intuitive judgments of justification?

In much the way that the organon offers a second-order sorting of first-order processes, our pretheoretic process typing might group together the processes that A&P recognizes. People ordinarily seem to type processes at a very coarse grain, like seeing, hearing, remembering, having a hunch, etc. (Jönsson 2013). Nevertheless, we do intuitively take some cases of seeing to confer less justification than others, so if these justifiedness intuitions are driven by our intuitive process typings, we must be using more fine-grained types, at least some of time (Jönsson 2013, p. 260, Olsson 2016, p. 190). Although Jönsson doesn't test for it, it is plausible that the folk are sensitive, in this finer-grained typing, to known parameters like distraction, bad lighting conditions, partial occlusion (this is the example Olsson 2016 uses), etc. Suppose further that the folk came to believe (correctly) that the visual processing of shape and the visual processing of color were independent of each other and (incorrectly) that one is much less reliable than the other: would they change their intuitive typings to such categories as seeing colors and seeing shapes? They might. In this way, A&P might be viewed as capturing not what the folk already implicitly think, but what they *would* think, were they suitably informed.

However, I don't want the success or failure of A&P to hang directly on folk intuitions about particular cases. Some reliabilists are staunch methodists: if any intuitions are essential support for an epistemological theory, it's intuitions about general epistemic principles (e.,g., 'justification must be truth-conducive'), not about particular cases (e.g., 'Norman the clairvoyant is unjustified'). But even if we're particularists enough to think that case-intuitions strongly constrain our epistemological theorizing, it would be a quite radical particularism that held every one of these intuitions to be indefeasible. Even if (rather

implausibly) our epistemic principles derived *all* their justification from case-intuitions, that's no reason why those principles, perhaps in conjunction with empirical facts, couldn't come back around to challenge *some* of those intuitions. I think most epistemologists are, reasonably, neither staunch methodists nor radical particularists, nor methodological purists of any sort: the best epistemological theory is the one that best accommodates all the data, including empirical facts, general epistemic principles, and particular case-intuitions (even if we disagree about the relative weighting of these data).

In the end, I won't take it to be a crippling objection to A&P if it yields surprising verdicts about particular cases, which the folk would *never* agree to—provided these verdicts can be given an adequate rationale. (We may see some of these in the next section.) A&P is intended as part of a normative theory about epistemic justification, not as part of a psychological theory of intuitive justification attributions.

5. Taking processes seriously

One reason why it's fruitful to pursue a theory of process individuation without worrying about environments is that process reliabilism is a conjunction of two separable views: one about the epistemic role of reliability, the other about the epistemic role of processes. The process part of process reliabilism emphasizes the claim that the justification of a belief is determined by the *psychological provenance* of that belief—of how likely an arbitrary belief *with that provenance* is to be true. This might well differ from the objective unconditional propensity of that belief to be true or its probability of being true conditional on other factors. Conflating the psychology and the organon is symptomatic of one particular failure to fully appreciate this essential aspect of process reliabilism, but there are others, to which I now turn.

5.1 Content and mere content

The role of contents in A&P is nuanced in a way that I've only hinted at so far. I've insisted that differences of mere content don't ipso facto make for differences of process. It's not that content *can't* make a difference, only that a difference in content doesn't automatically make for a difference in process. It might well be that some contents (e.g., *face*, *word*, *social exchange*) trigger the execution of specialized algorithms or subroutines, and nothing in A&P precludes this.

What A&P does preclude is that contents serve as parameters *when there are no law-like psychological generalizations ranging over those contents*. This has important consequences.

Suppose someone uses a single algorithm for identifying birds, but unbeknownst to her, she's much more reliable with some species than with others. A&P requires us to say that in at least some such cases, all the bird beliefs are the result of the same process, because neither reliability, nor species, nor the presence of lookalikes, etc. counts as a parameter: none of these is a psychological variable that systematically affects processing. So I'll have to say that the beliefs are equally justified.

Some reliabilists think this is the wrong answer.[20] I have some initial sympathy for this opposition, but I think if we focus on the processes involved, we will see that we really don't want to cleave off the reliable operations of an unreliable process or the unreliable operations of a reliable one. Suppose that to multiply *m* and *n*, I simply *add m* and *n*. I "multiply" 2 x 2 using this method and get the right answer. Am I justified? Of course not. This is exactly the sort of reason why we wanted to link justification to the reliability of *processes* in the first place, rather than, say, simply to the probability that the belief is true. A&P implies that my belief from this algorithm that 2 x 2 is 4, is just as unjustified as the

[20] Michael Bishop and Klemens Kappel have said so in conversation. Beebe 2004 is committed to these beliefs having different degrees of justification.

other beliefs that result from it. This is clearly the proper verdict. Similarly, if my bird identification algorithm is a mostly good (reliable) one, then the resulting beliefs should be highly justified even regarding birds that the algorithm gets wrong—-even though it does so consistently; if it's a mostly bad algorithm, it shouldn't confer much justification even for those birds it happens to get consistently right.

In real-world cases, persistent reliability differences will often involve parameter differences. I might use seven diagnostic features to identify bluejays and only two to identify something as a crow. This—seven diagnostic features as opposed to two—is clearly a difference in parameter values, one that makes for a difference in reliability, although the reliability difference is determined by much more than just crows and blue jays. Alternatively, I might have a highly specific recognition template for blue jays, in the sense that, for a given extent of similarity space, I have a large number of templates, with relatively sharp tuning curves; at the same time, my "crow" template might be very broadly tuned and the only one in the neighborhood. These factors—neighborhood template density and tuning curves—have already been recognized as legitimate parameters, and differences in their values will typically bring about different degrees of reliability. In these sorts of cases, A&P allows for epistemic differences, but because of genuinely different processes, not because of differences in the reliability of particular uses of a single process.

What about two cognizers who are otherwise exactly alike, but in response to a particular stimulus, one judges that it's a bird, and the other judges that it's a pin footed sparrow? Shouldn't the first agent be more justified than the second? Yes, if everything else is equal, because the first is making a basic- or entry-level (see Rosch 1978) judgment, while the second is making a subordinate-level judgment. Surely there are psychological laws concerning subordinate- and basic-level categorization and concerning levels of specificity of

detectors. Thus, these factors are genuine parameters and can be utilized as such by A&P. Again, it's general parameters, rather than particular contents, that are doing the work.[21]

It's not obvious that my official definition of parameters from sections 2.2 and 2.3 above allows this response. The language and examples there made it seem as if only independent variables would count as parameters, since only differences among these will *cause* systematic differences in processing. Independent of the epistemological ramifications, it seems unprincipled to prohibit the cognitive process typing from taking output parameters into account in the same way it does input parameters. Therefore, I want to be explicit that a parameter is any psychological variable, differences in the value of which systematically cause *or constitute* differences in processing. 'Systematic' will still be understood in terms of psychological laws. Particular contents are still not parameters, because they fail the systematicity requirement. Truth, reliability, and the like still aren't parameters because they don't cause or constitute differences in *processing*. (This is important for reasons discussed in section 2.5.) Such factors well might affect the contents of the psychological states, but there aren't laws about how systems respond to true inputs (in contrast with, say, degraded or ambiguous inputs). And even though there are laws about how various factors affect reliability, these laws describe the *results* of processing, rather than processing itself; reliability isn't an *output* of a cognitive algorithm, even though of course some algorithms are reliable.

5.2 Organon and psychology, revisited

This basic idea generalizes in important ways. To give just one illustrative example, we might have initially thought it a condition of adequacy on any epistemological theory that

[21] Once more, these complications will tend to be unnecessary in real-world cases, since there are probably already separate algorithms for generic vs. specific visual categorization (Marsolek 1999). Still, it is important to allow that even without differences of algorithm, we could have distinct processes if, e.g., lighting conditions affected visual color judgments before it affected shape judgments, or if it affected subordinate or particular judgments before category judgments.

it classify as justified any beliefs that result from inferences that conform to *modus ponens*, and that it classify as unjustified any beliefs that result from inferences that conform to affirming the consequent. For the kinds of reasons just seen, I think this turns out to be wrong, and I think it speaks strongly in favor of A&P that it brings this surprising result to light.

I mentioned above that the organonic rules don't line up with the cognitive processes. Thus, somewhat surprisingly, the principles of epistemic evaluation derived directly from logical form (or from the organonic canon more generally) can serve only as rules of thumb, not as exceptionless generalizations. This means that inferences conforming to affirming the consequent aren't *necessarily* bad, even in completely standard environments.

Suppose, for example, that when confronted with a putatively deductive argument we quickly and unconsciously try to imagine a consistent model where the premises are true and the conclusion false. If we haven't succeeded in doing so after a few tenths of a second or so of trying, we judge the argument to be valid.[22] Such an algorithm could be used in evaluating the arguments of others, as well as in deciding what inferences to draw from premises we already hold. The number of premises and amount of time spent searching for models are among the obvious parameters, but the hypothesized algorithm is insensitive to logical form, which means that logical form can't make a difference to processing and thus can't be a parameter. Thus, fixing a set of parameters, some instances of the resulting process will conform to *modus ponens* while others will conform to affirming the consequent. In that case, A&P assigns the two conclusions the same degree of justification: the affirming-the-consequent-conforming inference token is just as justified as the *modus-ponens*-conforming inference token.

[22] This is a toy theory, oversimplified for the sake of illustration but loosely inspired by the mental models theory of deductive reasoning, for a recent overview of which, see Johnson-Laird 2012.

This is a striking result and might even seem like a refutation of A&P. In fact, I think it's the correct result, and it's to A&P's credit that it enabled the discovery of this surprising fact about justification.

Solving logic problems in this intuitive way is surely inferior to explicitly learning, remembering, and applying the rules of logic. But whether the intuitive process is pretty good (/reliable), terrible (/completely unreliable), or somewhere in between is largely a matter of how successful the agent is at thinking up counterexamples when they do exist. This process might therefore turn out to be highly reliable, in which case very few instances of affirming the consequent would slip through. But when they do, they're as justified as are the instances of *modus ponens* the process gets right.

That's the verdict that A&P forces on us, anyhow. On reflection, it seems to be the right view. For the same reasons that we want to allow for the possibility of justified but false beliefs, we should allow for justified conclusions that result from invalid inferences. If the process is an unreliable one, then it's just a spot of good luck when the agent uses it to draw a conclusion that happens to conform with *modus ponens*. If it's a reliable one, then it's just a spot of bad luck when she uses it to draw a conclusion that conforms with affirming the consequent. The good luck just mentioned doesn't make her any less unjustified, and the bad luck doesn't make her any less justified.

6. Case studies

Providing a theory of process types does far more for reliabilism than to simply stave off a famous objection. A detailed theory puts process reliabilism in a position to do some real positive work in epistemology. The final stage of my defense of A&P involves showing that it can do such work. I briefly examine two case studies—extant epistemological problems that reliabilism can, now that A&P provides a principled way of individuating

processes, shed light on. This not only illustrates the fruitfulness of the theory, but it further demonstrates the principled nature of the view: A&P, far from letting us cherry-pick processes in a self-serving way, actually forces certain process typings on us.

To this end, I need to first make explicit some assumptions that flesh out A&P in more detail.

6.1 Process and parameter internalism

I noted in section 2.1 that what processes cognitive psychology delivers will depend in part on the methodological and general substantive presumptions of the field. The dominant theoretical framework in cognitive psychology for the last half century or so has been a representationalist, information processing framework, one that takes psychological states to be internal representations and psychological processes to be sequences or sets of transitions among these psychological states. This view seems to be committed to a kind of "***process internalism***": psychological processes are entirely internal to the cognitive agent. This is a controversial view. Certain versions of behaviorist, enactive (Noë 2004), extended (Clark 2008), and Gibsonian psychologies, by contrast, might think of psychological states and/or processes as reaching partly out into the environment. Some philosophers think that some factive mental states, like knowing (Williamson 2000) and seeing (Martin 2002), are psychological states in the relevant sense, even though they aren't internal to the individual. Still others think that some cognitive processes, like testimony, and perhaps other social processes, span more than one individual (Goldberg 2010).

Despite these controversies, let's suppose process internalism is true. I don't have space to defend it here, but my interest in it for the present purposes is that it enables the further articulation of a detailed version of A&P. As mentioned above (section 3, item 2), different theorists can flesh this part out in different ways, even while embracing the core of A&P, given in section 2.4.

So far I have been carefully neutral about the nature of parameters. They are *psychological variables*, but that doesn't necessarily mean that they are variables that are also psychological states. They might, instead, include some of the external dependent and independent variables just mentioned. Partly as a way of expanding process internalism, but largely for the sake of illustration, let me also embrace "parameter internalism": all parameters are psychological states, events, or properties and are thus, in concert with process internalism, internal to the agent. This addendum imposes considerable constraints on cognitive process individuation. For example, we can't really count vision in lighting conditions *C* as a psychological type, because lighting conditions aren't a *psychological* variable in the now relevant sense. It's not how much light is out there illuminating the distal stimulus; it's how much more or less information about that stimulus becomes available to the organism as the causal consequence of the lighting (or what *kinds* of information—reductions in light intensity might alter color discriminations before affecting shape identification, etc.). If an external variable has a systematic effect on processing, then there will be some proximal (i.e., internal) variable by which it has this effect, although it's not always obvious what that proximal variable will be in every case. It is for this reason that I was careful above to talk about vision in good lighting conditions "*involving*" a different process from vision in bad light, rather than vision in good lighting conditions *being* a different process.

For the following, I will be supposing that process internalism and parameter internalism are true.

6.2 Hallucination

Indicator reliabilism (e.g., Alston 1988) differs from process reliabilism in holding that justification is a function of the ground on which the belief is based and the extent to which that ground reliably indicates the truth of the belief. (Reliable indication can be cashed

out in terms of simple objective conditional probabilities, as Alston does, or something more sophisticated; I'll focus on the simple version for purposes of illustration.) Grounds, for Alston, are the agent's evidence, or what the agent takes her evidence to be, or perhaps the causally relevant elements of these. Grounds are functionally analogous to premises and the beliefs based on them functionally analogous to conclusions (in some cases, they literally are premises and conclusions). The "argument" from ground to belief is a good one just in case the ground is a reliable indicator of the belief.

Process and indicator reliabilism pull apart in interesting ways. The former has at least a prima facie advantage over the latter regarding barn facade cases. Henry is looking a real barn but doesn't know he's in a region where nearly everything that looks like a barn is a fake. The standard intuitive response to this case is that Henry is Gettiered: he doesn't know, but he is still justified (Goldman 1976). Indicator reliabilism seems to imply otherwise, since the environment—apparently by stipulation—is one where barn experiences don't reliably indicate the presence of barns.[23] Process reliabilism, on the other hand, easily gets the right result: the process by which Henry comes to believe there's a barn is one that's used for many other, non-barn beliefs as well, and thus, even if reliability is determined by the environment of use (see section 1 above) the reliability of this process is hardly dented at all by the existence of barn facades. Whatever exactly the process is—something like shape-based visual recognition of artifacts, operating within parameters determined by the lighting conditions, current psychological set, matching template tuning curves and densities, etc.—it is very clearly a process that is used for many things other than barns: silos and tractors and sheds and train stations, probably desks and tables and staplers as well. So, in contrast with

[23] Can an indicator reliabilist say something clever about environments and conditional probabilities to avoid the counterintuitive result? Maybe (although I suspect we could build all that into our description of fake-barn-environment and start all over). I was careful to claim this merely as a *prima facie* advantage for process reliabilism; the purpose here is to illustrate the differences between the two views, not to put one of them to rest. What matters is that the indicator reliabilist has to respond to this kind of case very differently from how the process reliabilist does.

indicator reliabilism, the only way that the presence of barn facades could undermine Henry's justification on process reliabilism is by having a very *general* and therefore highly attenuated effect on the reliability of a rather general psychological process. What little undermining there is would apply to all the outputs of the process, not just the ones about barns.

The situation is reversed for the case of hallucination: here it is indicator reliabilism that has the prima facie advantage. The fact that I'm hallucinating just this once does very little to undermine the reliable indication relation between my experience and belief. Thus, hallucination beliefs (i.e., putatively perceptual beliefs that are formed during or as a result of hallucination) are still normally justified.[24] This—at least by most epistemologists' lights—is the intuitively correct verdict. But it's a verdict that process reliabilism with A&P can't get, because hallucination is going to at least sometimes involve a very different and much less reliable process than normal perception, even though the experiential element is indistinguishable.

This intuition, that hallucination beliefs remain justified, is surely not only had by internalists; many reliabilists have modified their theories in complicated ways in order to allow victims of a Cartesian demon to be justified when they believe as we do in accordance with experiences like ours (see the works cited in section 1, above). A&P's verdict about hallucination, therefore, is a potential problem, one that I take quite seriously.[25] I don't want to lose the support of reliabilists who might otherwise endorse A&P. I think I can turn it into a virtue of the theory, but it will take some explanation.

[24] The standard understanding of hallucination—the subject has a perception-like experience as of an object, but there is no such object that the subject thereby perceives—is adequate for the present discussion. For a more nuanced discussion, see Macpherson and Batty (2016).

[25] For what it's worth, I myself found this implication of A&P initially appalling and was considering abandoning the current project over it. (Similarly, for what *that's* worth, with the implication, discussed in section 5.2, that affirming the consequent can be as justified as *modus ponens*.)

First off, I want to emphasize that the fact that A&P forces *any* epistemological verdicts on us, is a mark in favor of it. A solution to the generality problem needs to take the typing of the relevant processes out of our hands, so we can't just cherry-pick process types to get the cases right, and this is an illustration of how A&P does that (see point 2 in section 3, above). That mark in favor might be overridden if the verdict is simply unacceptable. I aim to show that it's not, or at least shouldn't be, once we think more about what's going on in cases of hallucination.

Notice that when A&P is augmented with process internalism and parameter internalism, it allows *some* hallucination beliefs to result from the same process as normal perception. If a demon or mad neuroscientist were to stimulate the sense organs or earliest brain areas in just the way distal objects would, the resulting process would be the same as one involved in normal, veridical, reliable perception. Process and parameter internalism prohibit the presence or absence of distal stimuli from making a difference to the process type, so these hallucination beliefs will be as justified as their perceptual counterparts.[26] (Such hallucination needn't be widespread enough to significantly reduce the reliability of the process, but even if it is, the diminished epistemic status will apply to all outputs of that process, perception and hallucination alike.) Justified hallucination beliefs are possible even on the kind of reliabilist view that withholds justification from demon world victims.

On the other hand, however, I would understand something like Macbeth's dagger experience as originating in conceptual and affective states and involving a *highly* top-down kind of processing quite unlike what happens in ordinary perception.[27] Such an endogenous

[26] Although process and parameter internalism are assumed for the sake of the current discussion, an epistemological disjunctivist (one who holds that hallucination ipso facto provides less justification than veridical perception) would surely reject one of them, probably both. In fact, embracing A&P while rejecting process internalism might *entail* something like epistemological disjunctivism.

[27] There are active debates, of course, about how much, if any, perceptual processing is top-down, or conceptually driven (Raftopoulos 2005, Hohwy 2013, Clark 2013), but I think that proponents of even the most radical predictive processing view, according to which perception is just "controlled hallucination" (Clark 2015) will accept that there's a difference of cognitive process here (hence "controlled").

process would involve a very different algorithm from normal perception, one that's surely much less reliable. So A&P has to say that at least *these* sorts of hallucination beliefs are unjustified.

I think reliabilists should not be bothered by this. A&P does require us to reject the principle that the justification of a perceptual belief is determined by the intrinsic nature of the accompanying perceptual experience. But we already had reasons to reject that principle. Suppose a novice prospector's desire to find gold makes it look to him as if the yellow rock in his pan is gold (Markie 2005). This case of "wishful seeing" seems epistemically flawed in much the way that wishful thinking is, and intuitively, the prospector's perceptual belief is unjustified, even though a qualitatively identical experience would have been enough to justify an expert who's not engaged in wishful seeing. Or suppose that Jill thinks—for no good reason—that Jack is angry with her, and this causes her to see his (neutral) facial expression as angry. She believes on the basis of this visual experience that Jack is angry (Siegel 2012). Jill's experience doesn't seem to have the epistemic force that qualitatively similar experiences would normally have. In these cases, it's intuitively appealing to claim that this is true because of the bad, endogenous process by which they're produced. Were phenomenally identical experiences to be produced in the normal, exogenous way, they'd produce justified beliefs, but the etiology of the experience seems, in these cases, to ruin the normal justificatory force. These are cases of illusion, rather than hallucination, but they are cases where phenomenally identical experiences differ in their epistemic power, and they do so at least partly because of their endogenous origin.

If wishful seeing, and seeing because you already believe it, can render an illusory experience epistemically inert (i.e., incapable of conferring justification), why couldn't, say, terror, guilt, and madness do the same to a hallucinatory experience? A lot more has gone wrong with Macbeth than with Jill, after all. If we have already agreed that the causal history

of an experience can affect that experience's ability to justify beliefs, then there shouldn't be any automatic resistance to the idea that some hallucinatory beliefs might be unjustified, especially when we remember that it is not the mere lack of veridicality that is blocking justification, but the fact that the experience is endogenously generated in a way that makes it very likely to be false. No, the hallucinator doesn't realize she's hallucinating, but the wishful seer doesn't realize that that's what she's doing either.

Not everyone agrees that wishful seeing could ruin the justification of a perceptual belief (see Fumerton 2013, Heumer 2013), but a growing number of epistemologists do, or at least take it quite seriously, and apparently on intuitive grounds.[28] Many of these authors, including the originators of the prospector and angry-looking-Jack cases, are epistemological *internalists*. Reliabilists of all people ought to welcome these intuitive verdicts. In the end, I think that if Jill is unjustified in believing that Jack is angry, then there's no reason to deny that Macbeth is unjustified in believing there's a dagger in front of him.

Up until now, the epistemology of hallucination has been almost exclusively dominated by two frameworks. "Traditionalism" holds that hallucination beliefs are (always) just as prima facie justified as the corresponding veridical perceptual belief. Epistemological disjunctivism, on the other hand, holds that a hallucination belief is—simply in virtue of depending on hallucination—less justified than the corresponding veridical perceptual belief.[29] A&P allows (forces, actually) a third view, somewhat in between these two: some hallucination beliefs are justified, and some are unjustified. Against disjunctivism, the mere fact that one is hallucinating does not by itself undermine justification. But against

[28] See, e.g., Brogaard 2013, McGrath 2013, Vahid 2014, Vance 2014, Ghijsen 2016, Teng 2016, Munton 2017.
[29] *Metaphysical* disjunctivism holds, roughly, that the mental state you are in while hallucinating is different in kind from the one you're in when you're veridically perceiving. See Haddock & Macpherson (2008) for more on this view. This view and epistemological disjunctivism are independent. For defenses of epistemological disjunctivism (though they don't always call it that), see McDowell 1998, Millar 2011, Pritchard 2012, Byrne 2016, Schellenberg 2016.

traditionalism, the intrinsic character of the experience doesn't determine justification either. A&P thereby opens up an interesting, plausible, and I think important position in the epistemology of hallucination, one that to my knowledge hasn't been defended in print before, certainly not in any detail.[30]

6.3 Cognitive penetration of perception

The prospector and angry-looking-Jack cases from the last section figure into a larger epistemological debate that has gained prominence in recent years. The "cognitive penetration of perception" is the empirical hypothesis that our beliefs, hopes, fears, expectations, etc. (in this context called "cognitive" states) might be able to directly influence perception. This hypothesis is the center of a lively controversy in the philosophy of cognitive science (Fodor 1983, Pylyshyn 2003, Raftopoulos 2009, Macpherson 2012, Lupyan and Ward 2013, Stokes 2013, Vetter and Newen 2014, Lupyan 2015, Firestone and Scholl 2016, Wu 2017). The question in epistemology concerns the hypothetical epistemological consequences: *if* cases like the prospector or angry-looking-Jack were to occur, would the epistemic potency of the resulting experiences be in fact reduced, and if so, why?

These stock cases might give the impression that the epistemic implications of cognitive penetration could only be negative. Yet proponents of the empirical thesis of

[30] I know of two authors who come close to this view. Susanna Siegel (2017) has an overall theory about the justificatory power of perceptual experience that (although quite different otherwise) is highly congenial to something very much like the current view about hallucination; however, I can't find her addressing the epistemology of hallucination anywhere in the book. Tyler Burge briefly mentions these issues in a footnote (note 24) in his (2003). This is a highly compressed discussion embedded in a complex and difficult paper. Insofar as I understand what's going on in that footnote, his view and mine agree roughly about which hallucination beliefs are justified, although for very different reasons. He seems to think that as long as an experience is genuinely perceptual and reliable, it confers justification, even if its putative object doesn't exist. However, genuine hallucinations (including what I'm calling endogenous hallucinations), which he contrasts with cases of mere referential illusion (what I'm counting as hallucination more generally), don't count as *perceptual* and therefore don't inherit the justification-conferring power from perceptual systems. Thus, his distinction between justified and unjustified hallucinations is based on function, not process reliability. One consequence, I think, is that this difference would have to be all-or-nothing on his view, while on mine it will come in degrees. Furthermore, my contrast between Cartesian and Macbeth-style hallucinations is intended merely to illustrate extremes of reliability differences; the endogenous/exogenous distinction doesn't carry any epistemic weight in and of itself, provided there might be reliable endogenous processes and unreliable exogenous ones.

cognitive penetration have traditionally argued that such top-down processing exists in part because it makes us *better* at perceiving our environment (Churchland 1988, Clark 2013). If that's right, then penetration shouldn't always be a bad thing. If my irrational fear of snakes makes me better at spotting the snakes that are really there—without increasing my rate of false positives—then my perceptual beliefs about snakes should still be perfectly well justified, despite the penetration (Lyons 2011). Thus, there may be epistemically good as well as bad cases of cognitive penetration. Process reliabilism offers a pretty obvious and attractive account of which ones are which: cognitive penetration is epistemically bad when it makes us worse at getting things right, good or innocuous otherwise (Lyons 2016a). This offers a natural explanation of the prospector, angry-looking-Jack, and snake cases, as well as the differences among them, and fits the whole issue into a larger, independently motivated epistemological framework.

But can reliabilists really get away with this so easily? Chris Tucker (2014) argues that we can't; in fact, he argues, reliabilism has a worse time making sense of cognitive penetration than his preferred internalist dogmatism does.

Tucker has two arguments designed to show that reliabilism can't offer a principled account of why wishful seeing is less justified than normal seeing. One of them trades pretty explicitly on the individuation of processes. Imagine a perceptual process that's very unreliable when it's being influenced by desires but quite reliable otherwise, the former instances being sufficiently rare that the overall reliability of the process is still high. The reliabilist, Tucker says, will have to claim that the wishful seeing that results from this process yields justified beliefs. But this seems wrong.

There's a ready answer to this, already developed above for completely unrelated purposes: desire is a psychological variable that has systematic effects on processing and is thus a parameter. Beliefs that are highly influenced by desires are therefore the results of

different processes from those that are only moderately so influenced, or not influenced at all. Thus, Tucker's supposition is incoherent according to A&P; if desire is a parameter, desire-influenced perception *can't* involve the same process as desire-independent perception.[31] And it's the reliability of the desire-influenced perceptual process that determines the degree of justification, thus, presumably, giving us the intuitively correct answer after all for the cases of wishful seeing.[32],[33]

Still, there's a more general challenge here, which is that reliabilism needs to show that it has the resources to give a detailed solution to the epistemic problem of cognitive penetration: to be able to say not only why cognitive penetration results in epistemic downgrade in those cases where it does, but also why it doesn't result in epistemic downgrade in other cases. This can only be done in the context of a worked out theory of cognitive processes. But it will also require a more detailed empirical treatment of the algorithms and parameters involved in the actual cases. Although we aren't in a position to specify the relevant algorithms in detail, we can consider different families of algorithms by looking at the reliability difference that various sub-algorithms (algorithmic subroutines) would make to existing algorithms.

---

[31] Desire is *at least* a parameter. In principle, desire could trigger the use of a different algorithm. Either way we have a difference in process. Tucker is, of course, well aware that his objection requires pinning a particular process typing on the reliabilist and thus butts up against the generality problem. Authors in this position, writing against reliabilism, almost never explicitly address any of the several published solutions to the generality problem to see if their objections work against detailed versions of reliabilism that have actually been defended. The one solution Tucker briefly mentions (Beebe 2004) can offer a response similar to mine, by appeal to statistically relevant factors (see section 2.5 above). In this case, because desire is both a parameter and a statistically relevant factor, our theories give the same result.

[32] Tucker claims that an "input-restricted reliabilism," which makes any difference in inputs yield a difference of process, falls prey to familiar objections to indicator reliabilism. What I'm urging is quite different; not every difference of input makes for different processes, only differences of *parameters*.

[33] I mentioned that Tucker has two arguments here. The other one is that it's possible to have a wishful seeing process that's made reliable by the intervention of a benevolent demon. Since such beliefs are intuitively unjustified even though reliably formed, this is a problem for reliabilism. This argument is a strange choice for showing that reliabilism has problems handling cognitive penetration, since, for one thing, it's only a problem if the reliabilist endorses a highly optional and rather unpopular view about the environment of reliability assessment (recall the discussion from section 1). For another thing, the objection has nothing to do with cognitive penetration; it's an old objection (going back to Goldman 1979) and quite general in its scope. Even if it shows that reliabilism has a problem, it doesn't show that it has a problem *with cognitive penetration*.

Here, for example, are some possible ways in which cognitive penetration might come about:

(a) perceptual categorization is achieved by weighted summation of diagnostic features, and the penetration results in giving more weight to the confirmatory features and less to the disconfirmatory features,

(b) categorization is realized by the firing of a detector, and the firing thresholds for the corresponding detectors are lowered in response to the penetrating cognitive states,

(c) feature-based attentional modulation simply leads to selection and more thorough processing of stimuli that are initially perceptually similar to items that are conceptually salient,

(d) conceptual salience of certain stimuli leads to suppression of processing for competitor stimuli,

(e) perceptual imagery is generated, which interacts with perceptual processing and nudges it in the direction of the imagery.[34]

These are five different sub-algorithms which, when added to or incorporated into existing algorithms, would result in cognitive penetration of perception. The first two sub-algorithms seem likely to detract from reliability (and thus reduce justification), while the next two seem likely not to (the fifth would need to be spelled out in more detail before its reliability effects would be clear). To put it in terms from signal detection theory, (a) and (b) are mechanisms that would normally lower *criterion* (aka *c*), while (c) and (d) would raise *sensitivity* (aka *d'*), meaning that, although all would result in a greater number of categorizations in line with

[34] These mechanisms I'm suggesting are speculative, although all are based on or inspired by actual proposals (not all of which are concerned with cognitive penetration). For instance, (c) has been proposed by Zhou and Desimone 2011, Bichot et al. 2015, among others; (e) has been proposed by Macpherson 2012. (a) is similar in spirit to the well-known confirmation bias (Evans 1989, Nickerson 1998), as (b) is to work by Sridharan et al. 2017, Luo & Maunsell 2015, 2018 (although much of this work is concerned with spatial attention, rather than feature-based attention), and (d) to Shulman et al. 2007, Larson and Lee 2013, Chelazzi et al. 1998, 2001, Desimone and Duncan 1995.

expectations (or fears, or wishes, or whatever is doing the penetrating), (a) and (b) would be accompanied by an increase in false positives, while (c) and (d) would not, but would reduce the number of false negatives, or misses.[35] *How much* the mechanisms would raise sensitivity (/reduce misses) or lower criterion (/increase false positives) would depend on the values of the parameters: how many diagnostic features are now being significantly employed, whether the reweighting was random or driven toward a particular result, how much the threshold is lowered, how similar and how numerous the suppressed competitors were, how much more processing is now allocated to the selected stimuli, etc.

Our pretheoretical intuitions about fictional cognitive penetration cases couldn't plausibly be driven by a detailed understanding of these sorts of mechanisms. But they could be driven by a second-order sorting of first-order processes (see section 4.2). The standard examples contain three salient though not always explicit features: (i) the nature of the penetrator, (ii) the truth-value of the perceptual belief, and (iii) an often implicit claim about whether the process now used is one that likely would have led the agent astray not just on this particular occasion but on many others as well, i.e., whether this process produces more false positives than normal perception. ((iii) is explicit in the snakes case but not in the others.) I suspect it's really the third, often implicit one here that's doing all the work generating our intuitive verdicts. I think we interpret the prospector and angry-Jack cases as involving some mechanism that, like (a) or (b), lowers criterion and reduces the reliability of the overall process, and we interpret the snakes case as involving a mechanism that, like (c) or (d), raises sensitivity without reducing the reliability of the overall process. If Jill's unjustified belief sensitized her to Jack's emotions in a way that reduced her false negatives

[35] Whatever the exact algorithms, there is growing evidence that we have at least two mechanisms for selectively attending, a pathway involving the superior colliculus that lowers criterion, and one involving V4 (and perhaps IT) that raises sensitivity (Chelazzi et al. 2001, Luo and Maunsell 2015, Sridharan et al. 2017), though see also Lovejoy and Krauzlis 2017.

without increasing her false positives, we'd think she's still justified—even on a given instance where she happened to be wrong.[36] Similarly for the prospector. Thus, it is still intuitions about processes, and process reliability, that are driving our epistemic intuitions about the cases.

If this is right, then A&P can account for our intuitive epistemic judgments about the hypothetical cases, while at the same time providing the framework for a principled treatment of whatever actual cases there might turn out to be.

Importantly, A&P prohibits the *epistemic status* of the penetrator from being a parameter (it's not a psychological variable that systematically affects processing), nor does it plausibly affect the algorithm; thus, according to A&P, the epistemic status of the penetrator cannot make a difference to process type or, consequently, to justification. This is exactly right. The reliabilist should insist that, even if it's a completely irrational belief that penetrates my snake perception, that perception might be epistemically impeccable if the penetration increases my sensitivity; and even if Jill's perceptual insensitivity is brought about by the well justified beliefs that Jack's mother just visited and that he's frequently angry after her visits, Jill's *perceptual* belief remains unjustified if these beliefs' penetration makes her perceptual process a less reliable one (see also Lyons 2011).

7. Conclusion

I've developed an algorithm-and-parameters (A&P) solution to the generality problem for reliabilism, in particular, to the problem of type-individuating cognitive processes. I've contrasted it with other solutions and other understandings of what processes might be. I've illustrated some of the ways in which the core proposal can be elaborated in light of

[36] Siegel (2013, p. 716) seems to agree that Jill's perceptual belief would be justified if it arose under such conditions.

independent, general cognitive scientific views, like process and parameter internalism. I've shown how an elaborated theory of processes can be applied to some specific epistemological problems.

The theory I've proposed delivers a unique type for every process token. It does this in a principled way, both in that the typing is theory-driven rather than ad hoc, and in that the typing is independently motivated: the typing latches on to (non-epistemic) joints in nature. A&P integrates with our best science, and it does so in a way that allows scientific findings to yield epistemological discoveries. It individuates processes at an appropriate level of generality to give us the epistemic distinctions we need; and it gives us good verdicts about actual and hypothetical cases.

These aren't always the verdicts that we would have endorsed without guidance from A&P. The verdicts concerning affirming the consequent (section 5.1) and hallucination (6.2) in particular are surprising. But far from an objection, this illustrates the two most important features of A&P: first, that it is sufficiently principled as to force epistemic verdicts on us, whether we like them or not; and second, that this makes A&P a source of epistemological discovery—something epistemological theories rarely aspire to. On reflection, and taking seriously the *process* part of process reliabilism, we can see that these initially counterintuitive results are quite plausible after all. Even in less dramatic cases, like cognitive penetration, we can see how having a detailed and principled theory of individuating processes can help us to solve outstanding problems in epistemology, problems that may have seemed unrelated to the generality problem. For this reason, even if one didn't care at all about the generality problem, one should still care about A&P.[37]

[37] Ancestors, versions, and/or parts of this paper have been presented at the Midwest Epistemology Workshop, Southeastern Epistemology Conference, Bled Philosophical Conference, Flemish Epistemology Workshop, University of Cardiff, University of Cambridge, and University of Glasgow. Thanks to those audiences, and especially to Ori Beck, Michael Bishop, Bob Beddor, Adam Carter, Jennifer Corns, Joseph Cruz, Julien Dutant, Branden Fitelson, Richard Fumerton, Harmen Ghijsen, Sanford Goldberg, Alvin Goldman, Peter Graham,

References

Alston, William P. 1988. "An Internalist Externalism." *Synthese* 74: 265-83.

——-. 1995. "How to Think About Reliability." *Philosophical Topics* 23: 1–29.

Beebe, James. 2004. "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis." *Nous* 38: 177-95.

Beddor, Bob. 2015. "Process Reliabilism's Troubles With Defeat." *Philosophical Quarterly* 65: 145-59.

Bergmann, Michael. 2006. *Justification Without Awareness: A Defense of Epistemic Externalism.* New York: Oxford University Press.

Bichot, Narcisse P., Matthew T. Heard, Ellen M. DeGennaro, and Robert Desimone. 2015. "A Source for Feature-Based Attention in the Prefrontal Cortex." *Neuron* 88: 832-44.

Biederman, Irving. 1987. "Recognition-By-Components: A Theory of Human Image Understanding." *Psychological Review* 94: 115-47.

Bishop, Michael A. 2010. "Why the Generality Problem is Everybody's Problem." *Philosophical Studies* 151: 285-98.

Braine, Martin. D. S., and David P. O'Brien, eds. 1998. *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brogaard, Berit. 2013. "Phenomenal Seemings and Sensible Dogmatism." In *Seemings and Justification*, ed. Chris Tucker, 270-289. Oxford: Oxford University Press.

Burge, Tyler. 2003. "Perceptual Entitlement." *Philosophy and Phenomenological Research* 67: 503-48.

David Henderson, Christopher Hill, Klemens Kappel, Christoph Kelp, Rae Langton, Adam Leite, Clayton Littlejohn, Jon Matheson, Kevin McCain, Matthew McGrath, Hugh Mellor, Andrew Moon, Orestis Palermos, Ted Poston, Wayne Riggs, Sven Rosenkranz, Thomas Senor, Mona Simion, Alessandra Tanesini, Timm Triplett, Chris Tucker, Chase Wrenn, Sarah Wright, and Dean Zimmerman. Thanks also to an editor and two anonymous referees for this journal.

Byrne, Alex. 2016. “The Epistemic Significance of Experience.” *Philosophical Studies* 173: 947–67.

Chelazzi, Leonardo, John Duncan, Earl K. Miller, and Robert Desimone. 1998. “Responses of Neurons in Inferior Temporal Cortex During Memory-Guided Visual Search.” *Journal of Neurophysiology* 80: 2918–40.

Chelazzi, Leonardo, Earl K. Miller, John Duncan, and Robert Desimone. 2001. “Responses of Neurons in Macaque Area V4 During Memory-Guided Visual Search.” *Cerebral Cortex* 11: 761-72.

Churchland, Paul M. 1988. “Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor.” *Philosophy of Science* 55: 167-87.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.

——-. 2013. “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” *Behavioral and Brain Sciences* 36: 181-204.

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

Collett, Thomas S., and Paul Graham. 2004. “Animal Navigation: Path Integration, Visual Landmarks and Cognitive Maps.” *Current Biology* 14: R475-R477.

Coltheart, Max. 2005. “Modeling Reading: The Dual-Route Approach.” In *The Science of Reading: A Handbook*. ed. M. J. Snowling and C. Hulme, 6-23. Oxford: Blackwell.

Comesaña, Juan. 2002. “The Diagonal and the Demon.” *Philosophical Studies* 110: 249–66.

——-. 2006. “A Well-Founded Solution to the Generality Problem.” *Philosophical Studies* 129: 27-47.

Conee, Earle, and Richard Feldman. 1998. “The Generality Problem for Reliabilism.” *Philosophical Studies* 89: 1–29.

Dehaene, Stanislav. 2011. *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.

Desimone, Robert, and John Duncan. 1995. "Neural Mechanisms of Selective Visual Attention." *Annual Review of Neuroscience* 18: 193-222.

Evans, Jonathan S. B. 1989. *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: Erlbaum.

Evans, Jonathan S. B., Julie L. Barston, and Paul Pollard. 1983. "On the Conflict Between Logic and Belief in Syllogistic Reasoning." *Memory and Cognition* 11: 285-306.

Evans, Jonathan S. B., and Keith E. Frankish. 2009. *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, Jonathan S. B., and Keith E. Stanovich. 2013. "Dual-Process Theories of Higher Cognition: Advancing the Debate." *Perspectives on Psychological Science* 8: 223-41.

Farah, Martha J. 2004. *Visual Agnosia*. Cambridge, MA: MIT Press.

Feigenson, Lisa, Stanislas Dehaene, and Elizabeth Spelke. 2004. "Core Systems of Number." *Trends in Cognitive Sciences* 8: 307-14.

Firestone, Chaz, and Brian J. Scholl. 2016. "Cognition Does Not Affect Perception: Evaluating the Evidence for 'Top-Down' Effects." *Behavioral and Brain Sciences* 39: 1-72.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, Jerry A., and Zenon W. Pylyshyn. 1988. "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition* 28: 3-71.

Forster, Kenneth I., and Susan M. Chambers. 1973. "Lexical Access and Naming Time." *Journal of Verbal Learning and Verbal Behavior* 12: 627-35.

Fumerton, Richard. 2013. "Siegel on the Epistemic Impact of 'Checkered' Experience." *Philosophical Studies* 162: 733-9.

Gallistel, Charles R. 2007. “Dead Reckoning, Cognitive Maps, Animal Navigation and the Representation of Space: An Introduction.” In *Robotics and Cognitive Approaches to Spatial Mapping,* ed. Margaret E. Jefferies and Wai-Kiang Yeap, 137-44. Springer, Berlin, Heidelberg.

Gerlach, Christian. 2009. “Category-Specificity in Visual Object Recognition.” *Cognition* 111: 281-301.

Ghijsen, Harmen. 2016. “The Real Epistemic Problem of Cognitive Penetration.” *Philosophical Studies* 173: 1457-75.

Goldberg, Sanford. 2010. *Relying on Others*. Oxford: Oxford University Press.

Goldman, Alvin I. 1979. “What is Justified Belief?” In *Justification and Knowledge*, ed. George Pappas. Dordrecht: Reidel.

——-. 1976. “Discrimination and Perceptual Knowledge.” *Journal of Philosophy* 73: 771-91.

——-. 1986. *Epistemology and Cognition*. Cambridge: Harvard UP.

Graham, Peter J. 2012. “Epistemic Entitlement.” *Noûs* 46: 449–82.

Graham, Peter J., and Jack C. Lyons. n.d. “Reliabilism and Epistemic Defeat.” Unpublished manuscript.

Haddock, Adrian, and Fiona Macpherson, eds. 2008. *Disjunctivism: Perception, Action, Knowledge*. Oxford: Oxford University Press.

Heller, Mark. 1995. “The Simple Solution to the Problem of Generality.” *Noûs* 29: 501-15.

Henderson, David, and Terence Horgan. 2006. “Transglobal Reliabilism.” *Croatian Journal of Philosophy* 6: 171–95.

Hohwy, Jakob. 2013. *The Predictive Mind.* Oxford: Oxford University Press.

Howard, Ian P. 2002. *Seeing in Depth, Vol. 1. Basic Mechanisms*. Toronto: University of Toronto Press.

Huemer, Michael. 2013. “Epistemological Asymmetries Between Belief and Experience.” *Philosophical Studies* 162: 741-8.

Johnson-Laird, Phillip N. 2001. “Mental Models and Deduction.” *Trends in Cognitive Science* 5: 434-42.

——-. 2012. “Inference with Mental Models.” *The Oxford Handbook of Thinking and Reasoning*, ed. Keith J. Holyoak and Robert G. Morrison, 134-45. Oxford: Oxford University Press.

Jönsson, Martin L. 2013. “A Reliabilism Built on Cognitive Convergence: An Empirically Grounded Solution to the Generality Problem.” *Episteme* 10: 241-68.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kappel, Klemens. 2006. “A Diagnosis and Resolution to the Generality Problem.” *Philosophical Studies* 127: 525-60.

Larson, Eric, and Adrian K.C. Lee. 2014. “Switching Auditory Attention Using Spatial and Non-Spatial Features Recruits Different Cortical Networks.” *Neuroimage* 84: 681-7.

Lovejoy, Lee P., and Richard J. Krauzlis. 2017. “Changes in Perceptual Sensitivity Related to Spatial Cues Depends on Subcortical Activity.” *Proceedings of the National Academy of Sciences* 114: 6122-6.

Luo, Thomas Zhihao, and John H. R. Maunsell. 2015. “Neuronal Modulations in Visual Cortex Are Associated With Only One of Multiple Components of Attention.” *Neuron* 86: 1182-8.

——-. 2018. “Attentional Changes in Either Criterion or Sensitivity Are Associated With Robust Modulations in Lateral Prefrontal Cortex.” *Neuron* 97: 1382-93.

Lupyan, Gary. 2015. “Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems Are Penetrable Systems.” *Review of Philosophy and Psychology* 6: 547-69.

Lupyan, Gary, and Emily J. Ward. 2013. “Language Can Boost Otherwise Unseen Objects Into Visual Awareness.” *Proceedings of the National Academy of Sciences* 110: 14196-201.

Lyons, Jack C. 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: Zombies, Modules, and the Problem of the External World*. New York: Oxford University Press.

——-. 2011. “Circularity, Reliability, and the Cognitive Penetrability of Perception.” *Philosophical Issues* 21: 289–311.

——-. 2013. “Should Reliabilists Be Worried About Demon Worlds?” *Philosophy and Phenomenological Research* 86: 1-40.

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

——-. 2016b. “Goldman on Evidence and Reliability”. In *Goldman and His Critics*, ed. Hilary Kornblith and Brian McLaughlin, 149-74. Malden, MA: Blackwell.

Macpherson, Fiona. 2012. “Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism.” *Philosophy and Phenomenological Research* 84: 24–62.

Macpherson, Fiona, and Clare Batty. 2016. “Redefining Illusion and Hallucination in Light of New Cases.” *Philosophical Issues* 26: 263-96.

Markie, Peter J. 2005. “The Mystery of Direct Perceptual Justification.” *Philosophical Studies* 126: 347-73.

——-. 2006. “Epistemically Appropriate Perceptual Belief.” *Noûs* 40: 118-42.

Marsolek, Chad J. 1999. “Dissociable Neural Subsystems Underlie Abstract and Specific Object Recognition.” *Psychological Science* 10: 111-18.

Martin, Michael G. F. 2002. “The Transparency of Experience.” *Mind & Language* 17: 376-425.

McDowell, John. 1998. “Criteria, Defeasibility, and Knowledge.” In *Meaning, Knowledge, and Reality*. Cambridge, MA: Harvard University Press.

McGrath, Matthew. 2013. “Phenomenal Conservatism and Cognitive Penetration: The Bad Basis Counterexamples.” In *Seemings and Justification*, ed. Chris Tucker, 225-247. Oxford: Oxford University Press.

Millar, Alan. 2011. “How Visual Perception Yields Reasons for Belief.” *Philosophical Issues*, 21: 332–51.

Mou, Yi, and Kristy vanMarle. 2014. “Two Core Systems of Numerical Representation in Infants.” *Developmental Review* 34: 1-25.

Müller, Martin, and Rüdiger Wehner. 2010. “Path Integration Provides a Scaffold for Landmark Learning in Desert Ants.” *Current Biology* 20: 1368-71.

Munton, Jessie. 2017. “Perceptual Skill and Social Structure.” *Philosophy and Phenomenological Research*.

Nickerson, Raymond S. 1998. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology* 2: 175-220.

Noë, Alva. 2004. *Action in Perception*. Cambridge, MA: MIT Press.

Nozick, Robert. 1981. *Philosophical Explanations*. Harvard University Press.

Olsson, Erik J. 2016. “A Naturalistic Approach to the Generality Problem.” In *Goldman and His Critics*, ed. Hilary Kornblith and Brian McLaughlin, 178-96. Malden, MA: Blackwell.

Peterson, Mary A., and Gillian Rhodes, eds. 2003. *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*. Oxford University Press.

Pollock, John. 1984. “Reliability and Justified Belief.” *Canadian Journal of Philosophy* 14: 103-14.

Pritchard, Duncan. 2012. *Epistemological Disjunctivism*. Oxford: Oxford University Press.

Pylyshyn, Zenon W. 1984. *Computation and Cognition*. Cambridge, MA: MIT Press.

——-. 2003. *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Raftopoulos, Athanassios. 2005. *Cognitive Penetrability of Perception: Attention, Action, Strategies, and Bottom-Up Constraints*. New York: Nova Science.

——-. 2009. *Cognition and Perception: How Do Psychology and Neural Science Inform Philosophy?* MIT Press.

Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*, ed. Eleanor Rosch and B. B. Lloyd, 27-48. Hillsdale, NJ: Erlbaum.

Schellenberg, Susanna. 2016. "Phenomenal Evidence and Factive Evidence." *Philosophical Studies* 173: 875–96.

Schneider, Walter, and Richard M. Shiffrin. 1977. "Controlled and Automatic Human Information Processing I: Detection, Search and Attention." *Psychological Review* 84: 1–66.

Schwarz, Norbert. 2002. "Situated Cognition and the Wisdom of Feelings: Cognitive Tuning." In *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence*, ed. Lisa Feldman Barrett and Peter Salovey, 144-66. New York, NY: Guilford Press.

Shastri, Lokendra, and Venkat Ajjanagadde. 1993. "From Simple Associations To Systematic Reasoning: A Connectionist Representation of Rules, Variables and Dynamic Bindings Using Temporal Synchrony." *Behavioral and Brain Sciences* 16: 417-94.

Shulman, Gordon L., Serguei V. Astafiev, Mark P. McAvoy, Giovanni d'Avossa, and Maurizio Corbetta. 2007. "Right TPJ Deactivation During Visual Search: Functional Significance and Support for a Filter Hypothesis." *Cerebral Cortex* 17: 2625-33.

Siegel, Susanna. 2012. "Cognitive Penetrability and Perceptual Justification." *Nous* 46: 201-22.

——-. 2013. "The Epistemic Impact of the Etiology of Experience." *Philosophical Studies* 162: 697-722.

——-. 2017. *The Rationality of Perception*. Oxford: OUP.

Smolensky, Paul. 1988. “On the Proper Treatment of Connectionism.” *Behavioral and Brain Sciences* 11: 1-23.

Sosa, Ernest. 1993. “Proper Functionalism and Virtue Epistemology.” *Nous* 27: 51-65.

——-. 2015. *Judgment and Agency*. Oxford: OUP.

Sridharan, Devarajan, Nicholas A. Steinmetz, Tirin Moore, and Eric I. Knudsen. 2017. “Does the Superior Colliculus Control Perceptual Sensitivity or Choice Bias During Attention? Evidence From a Multialternative Decision Framework.” *Journal of Neuroscience* 37: 480-511.

Strat, Thomas M., and Martin A. Fischler. 1991. “Natural Object Recognition: A Theoretical Framework and its Implementation.” *Proceedings of the International Joint Conference on Artificial Intelligence*.

Teng, Lu. 2016. “Cognitive Penetration, Imagining, and the Downgrade Thesis.” *Philosophical Topics* 44: 405-426.

Tucker, Chris. 2014. “If Dogmatists Have a Problem with Cognitive Penetration, You Do Too.” *Dialectica* 68 35-62.

Tye, Michael. 1991. *The Imagery Debate.* Cambridge, MA: MIT Press.

Vahid, Hamid. 2014. “Cognitive Penetration, the Downgrade Principle, and Extended Cognition.” *Philosophical Issues* 24: 439-59.

Vance, Jona. 2014. “Emotion and the New Epistemic Challenge from Cognitive Penetrability.” *Philosophical Studies* 169: 257-83.

Vetter, Petra, and Albert Newen. 2014. “Varieties of Cognitive Penetration in Visual Perception.” *Consciousness and Cognition* 27: 62-75.

Weatherson, Brian. 2012. “The Temporal Generality Problem.” *Logos & Episteme* 3: 117-22.

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

Witt, Jessica K., Dennis R. Proffitt, and William Epstein. 2005. “Tool Use Affects Perceived Distance, But Only When You Intend to Use it.” *Journal of Experimental Psychology: Human Perception and Performance* 31: 880-8.

Wunderlich, Mark E. 2003. “Vector Reliability: A New Approach to Epistemic Justification.” *Synthese* 136:, 237-62.

Zhou, Huihui, and Robert Desimone. 2011. “Feature-Based Attention in the Frontal Eye Field and Area V4 During Visual Search.” *Neuron* 70: 1205-17.