INTERNALISM AND IRRATIONALITY

Philosophy

Paper or Poster Session

Dr. Charles Wallis

California State University, Long Beach

1250 Bellflower Boulevard

McIntosh Humanities Building (MHB) 917

Long Beach, CA 90840-2408

cwallis@csulb.edu

562.985.4331

562.985.7135 (fax)

ABSTRACT FOR INTERNALISM AND IRRATIONALITY

A common counterexample to externalism and reliabilism asserts that it does not capture the sufficient conditions for justification since reliable processes can yield beliefs the that from the believer's own standpoint seem irrational. In this paper I argue that the internalist argument is based upon the false assumption that one's subjective perspective and one's psychological processes are easily and intuitively separable. I focus primarily upon the work of Lawrence BonJour. In his "Externalist Theories of Empirical Knowledge" as well as his *The Structure of Empirical Knowledge*, BonJour offers a series of counterexamples to reliabilist theories of knowledge. BonJour claims that these counterexamples illustrate a conclusive objection to reliabilist theories of knowledge. BonJour objects to reliabilism on the grounds that a person could have a belief counting as knowledge according to reliabilism, but which seems (epistemically) irrational in light of the person's own understanding of the evidence. BonJour's article and book have proven very influential. His counterexamples are arguable the most thorough treatment of this criticism in the philosophical literature and are often treated in the philosophical world as the final nails in the reliabilist's casket.

In this paper, I discuss BonJour's examples at length, arguing that the apparent strength of BonJour's objection is chimerical. I claim that the intuitive pull of BonJour's examples (and these examples generally) results, not from a weakness in reliabilism, but from rhetorical elements, vagaries, and false psychological presuppositions in BonJour's descriptions. A pure reliabilism like the one that I outline can handle these cases without recourse to the *ad hoc* moves and additional non-reliabilist clauses.

INTERNALISM AND IRRATIONALITY

1.) INTRODUCTION

In his "Externalist Theories of Empirical Knowledge" and his *The Structure of Empirical Knowledge*, Laurence BonJour offers a series of counterexamples to reliabilist theories of knowledge. BonJour claims that these counterexamples provide a conclusive objection to reliabilist theories of knowledgebecause they show that, according to reliabilism, a person could have a belief counting as knowledge, but which, nevertheless, seems (epistemically) irrational in light of person's own understanding of the evidence. BonJour's article and book have proven very influential. His counterexamples are arguable the most thorough treatment of this criticism of reliabilism in the philosophical literature and are often treated in the philosophical world as the final nails in the reliabilist's casket.

Interestingly, arguments like BonJour's counterexamples date back to the earliest statements of reliabilism (1976) and arguably do not refute the earliest and most influential version of reliabilism, Alvin Goldman's.[1] Alvin Goldman handles BonJour's counterexamples in his 1979 paper by adding a *no defeaters* clause to his definition of justification, and (in some cases) by interpreting the no defeaters clause to include defeaters that a person is *ex ante* justified in believing.

(Goldman 1986, pp.109-113) A person is *ex ante* justified in believing beliefs that they do not have, but which are permissible given a set of right J-rules and what the person does believe.[2]

If BonJour's examples do not refute Goldman's theory, wherein, then, lies the strength of BonJour's objection? Why does Goldman return to these irrationality examples with each new reformulation of this theory? (1976, 1979, 1986, 1991). I would suggest that the source of the strength of these counterexamples, and especially BonJour's counterexamples is twofold. First, BonJour's examples, as described by BonJour, look intuitively compelling. Second, the type of responses to BonJour offered by Goldman and others look *ad hoc*, and/or significantly out of step with reliabilism. Goldman's appeal to *ex ante* justification, for example, is made solely to deal with cases like those described by BonJour, and is never developed in Goldman's 1986 theory. The appeal to *ex ante* justification, therefore, looks *ad hoc*. Moreover, Goldman's appeal to a no defeaters clause evokes defeasability theory and not reliabilism to fend off these irrationality counterexamples.

In this paper, I discuss these irrationality examples at length, with an emphasis on BonJour. I claim that the intuitive pull of BonJour's examples (and these examples generally) results, not from a weakness in reliabilism, but from rhetorical elements, vagaries, and false psychological presuppositions in BonJour's and other's descriptions. A pure reliabilism like the one that I outline can handle these cases without recourse to the *ad hoc* moves and additional non-reliabilist clauses.

2.) OUTLINE FOR A RELIABILIST THEORY OF KNOWLEDGE

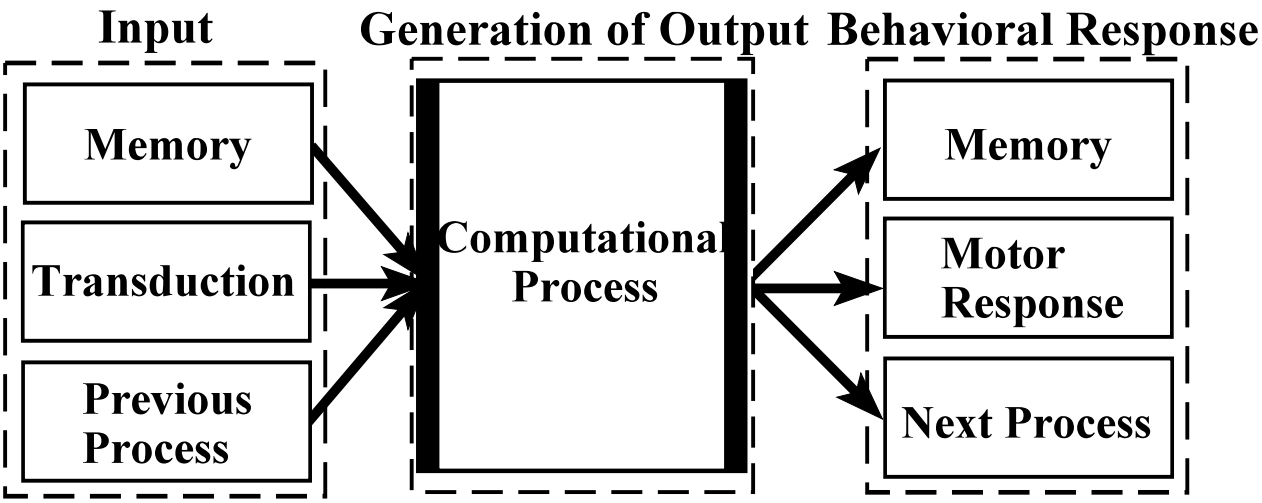
Anyone attempting to develop a reliabilist theory of knowledge must find solutions to three problems. I call these three problems *the problem of levels*, *the characterization problem*, and *the relevance class problem*.[3] The first problem, the problem of levels, concerns how (if?) and at what level one should set the minimum reliability of a process necessary for the process to generate

knowledge. The characterization problem concerns how to characterize cognitive processes in order to assess their reliability. The relevance class problem concerns how to delimit the class of uses of a process (class of situations) relevant to determining the reliability of a process.

As regards the problem of levels, BonJour assumes that the reliabilities of the processes he considers equal 1. While one might doubt that any or many real cognitive processes are perfectly reliable, such an assumption certainly does not beg the question against the reliabilist. I do not, therefore, consider solutions to the problem of levels in this paper. Rather, I claim that BonJour's examples, as he describes them, play upon vagaries and mischaracterizations as regards the characterization problem and the relevance class problem. Towards this end, I do outline a possible solution to both of the latter problems in some detail.

Before offering a sketch of my theory of knowledge, I wish to briefly present a basic outline of computational explanations in cognitive science. In order to fully understand my theory of knowledge, one must understand what I mean by a cognitive task and how the notion of a task figures in specifying a process and a task domain. The computational approach to cognitive science seeks to explain cognition in terms of three conceptual (though not necessarily temporally) distinct stages:

The first stage involves the input of representations, from other processes, from the transduction of physical parameters, from memory, or from some combination of the three. Transduction is thought of as the direct conversion of physical parameters in the environment into representations of those parameters. In the second stage, the system operates upon input, generating other representations as output. The model for this second stage is a computational model, though the style of computation varies widely. The representations generated in the second stage guide the system's performance in the final stage involving some motor response, input to memory, and/or input to other processes. For example, in explaining how I successfully pick-up objects, the computational



cognitive psychologist will talk of (at least) the reflected light input into my retina, the computational processes whereby I infer the nature and location of the object from retinal stimulations, and the motor response utilizing that object representation to perform the pick-up.

A well-specified task description includes the following three elements:[4] (1) A specification of input and output types in terms of an idealized target function that defines success conditions for performing the task. For instance, a planning program has requests and initial state descriptions as inputs and computes plans as outputs. A system computing the successor function has as input,

representations of numbers for which one wants the successor, and as output, representations of successors. (2) A specification of the nomic correlations (including statistical correlations) that underlie the system's performance of a task. For instance, David Marr (1976) discusses what he calls the "underlying physical assumptions" of human vision. Marr's assumptions amount to nomic generalizations about the task domain, and include the existence of surfaces, spatial continuity, and so on. (*Vision* pp. 44-51) (3) A specification of the relevant process by reference to the system's laws of operation within the domain, viewing these operations as a strategy or set of strategies for generating outputs from inputs relying upon certain nomic correlations.

The appropriate processes characterization is the one that reflects the cognizer's dispositions across all and only the cases that fall under the task specification. As a result, the process a cognizer is seen to instantiate represents the strategy that it employs in performing the task. The task domain --the set of relevant alternative situations--is determined by the specifications of the nomic correlations that underlie the system's performance of the task. The task specification determines what one understands to be the system's environment of habitual performance, which in turn determines the task domain. Note too, that the task, and hence the relevance class, will change should the system's environment undergo change or should the system come to inhabit different environments. One evaluates the effectiveness of the process (of the strategy) relative to one's best hypothesis as to the situations that the cognizer encounters in performing this new task.

My theory of knowledge provides one with unique process characterizations and task domain against the background of a well-specified task. However, formulations of task are based largely upon empirical investigation and argumentation as to, for instance, the system's laws of operation, the covering laws operant in the hypothesized domain, and the relative distribution of potential situations within the hypothesized domain. While one will not find widely disparate task

descriptions, and hence one will not find widely disparate hypotheses about relevance classes and process characterizations, there remains the possibility of disputes among informed individuals as to the exact formulation of a task.

I am now in the position to outline my reliabilist definition of knowledge. Outputs of a cognitive process count as knowledge iff the outputs are produced from inputs that count as knowledge, by some process, P, in performing some task, T, and the task-specific characterization of the process that produced the outputs has a high reliability in the task-specific relevance class.

One important reason why the theory presented here serves only as an outline is that a complete theory would specify what counts as knowledge for an input. Such a specification is made in part by the recursive nature of the definition. Indeed, in the places where inputs matter, it will be obvious how the recursive nature of the definition can supply them. However, a complete theory should deal with the possibility of innate knowledge, transduced inputs, and memory. Such an account, though too complex to present and defend here, is given in my dissertation, *Representation, Knowledge, and Structure in Computational Explanations in Cognitive Science*.

3.) RED BARNS

Before turning to BonJour's counterexamples, I wish to consider a different counterexample commonly forwarded against reliabilism. My purpose in considering this counterexample is to give the listener an example of how the above-outlined theory of knowledge applies to cases. It will also illustrate the importance of understanding the relevance class and characterization problems in discussing counterexamples to reliabilism.

The type of example that I want to consider is often attributed to Kripke as a criticism of Nozick based upon an example found in Goldman (1976), though Kripke has never published his version of the barn example. Such an objection might run as follows:[5]

Suppose that one is assigned the task of picking out barns from other building structures in the countryside. As a lifetime urbanite, you are not very reliable at distinguishing barns; but having seen many illustrations, you are very good at picking [out] red barns. It seems that in any given case one can know that there is a barn in the field by knowing that there is a red barn there, even though one's cognitive process is not generally reliable for the task at hand, i.e., even though it is not reliable for distinguishing barns from non-barns generally.

I think that the above description of the actual process of belief formation is psychologically inaccurate. Or, at least, one has a much more psychologically plausible model in Irving Biederman's theory of object recognition. I do not think that adopting Biederman's theory biases the case against the objector. Moreover, if Biederman's theory biases the case against such an objector, I am not sure why that would not further undermine the example. I will, therefore, introduce Biederman's theory and consider the as a example in light of that theory.

The theory I propose to use to frame the discussion, is a slightly modified version of Irving Biederman's "Recognition By Components" theory of object recognition.[6] I will not expound upon the virtues of Biederman's theory, except to say that it is one of the best documented theories I have come across. Moreover, it has a connectionist model that can do object recognition in real time (about 100 msec). While Biederman's theory is not the final theory for object recognition, it does represent the way in which cognitive scientists think about object recognition, and of cognition in general.

Biederman's theory supposes that the striate cortex codes sufficient information to detect certain types of volumetric primitives, which he calls "geons," and certain primitive spatial relations like *above*. Geons have the desirable property of being recognizable in almost any orientation and in many cases of partial occlusion. By assuming that object representations are computed by first

computing the geons and relations involved, Biederman's theory also has the nice feature that the number of detectable object/orientations is very large practically and infinite in the abstract.

Now, in the above example, one knows very little about country buildings. One does, however, know that large red buildings having a rectangular body and extended triangular roofs are barns. This last bit of knowledge proves important. In the terms of the Biederman-inspired object recognition system: One has a barn representation, S_b. One's barn representation gets tokened every time one sees a red building having a rectangular body and a triangular roof. Specifically, S_b gets tokened every time one's *rectangular* and extended *triangle* geons get tokened along with one's *red* and *above* (triangle above rectangle).

Having modified the above example in accordance with Biederman's theory, one should note that the example's task description is actually importantly ambiguous. One distinguishes barns from nonbarns. However, what is the final state (output) of the process? Does the process (task) end with each building that one judges? In other words, is one simply trying to identify buildings on a case by case basis? Or, is one trying to keep a running total of barns seen?

One must answer these last questions before one can answer questions like? Does one's tokening of S_b constitute knowledge? If the *rectangle-triangle-red-above* to S_b connection is reliable, then does one know that the building in question is a barn? What about one's general unreliability at picking out barns?

Let us consider the first case: Suppose one's task is to identify country buildings. One goes out to the country, and one's *rectangle* and *triangle* geons get tokened along with one's *red* and *above*, resulting in the tokening of an S_b.

We are now in a better position to evaluate the belief. One's barn recognizing behavior is broken down into several independent subprocesses. The verdict of one subprocess does not effect

the verdict of the other subprocesses. One of these subprocesses, $subprocess_1$, operates by taking the input of a *rectangle-triangle-red-above* to be a sufficient condition for recognizing a barn. $Subprocess_1$ reliably tokens an S_b when activated, and is activated only in those cases where the building is a red rectangular building with an extend triangle for a roof. Moreover, one can and does distinguish between cases where one uses $subprocess_1$ and other cases (though probably not under those descriptions).

The process characterization and relevance class must reflect one's disjoint strategy for identifying buildings. In other words, given the task description, the appropriate process characterization for the first case would take $subprocess_1$ as the process. The appropriate relevance class is the set of situations that activate $subprocess_1$. Since, $subprocess_1$ is reliable in all those cases where it is activated, one's tokening of S_b counts as knowledge. Given this task, the verdict of my approach coincides with the intuition that the example tries to cultivate.

Now, let us consider the second case. In the second case, the output is the number of barns. In this case, suppose, the output of the object recognition system serves as input for a counting process that keeps track of the number of S_bs tokened. Since one is not generally reliable at distinguishing barns from other buildings, other of one's subprocesses must not be reliable. For instance, suppose that one's only other subprocess for distinguishing barns from other country buildings is one that tokens a S_b whenever the input is not *rectangle-triangle-red-above*. This represents an unreliable strategy for the cases not covered by $subprocess_1$. So, in those cases where one uses $subprocess_2$, one does not know that the building in question is not a barn. Assuming that the environment of one's habitual performance includes enough instances of barns not activating $subprocess_1$, one's strategy of using both processes to determine the total number of barns will prove unreliable. In other words, one cannot know how many barns one has encountered, even though one

can know for some individual barns, that they are barns.

The difference in verdicts from the first case to the second case is explained by difference in our understanding of the processes and relevance classes associated with the two tasks. The original description of the example does not distinguish between cases. In fact, the original description contains elements of each task. On the one hand, it asserts that one is not very reliable at distinguishing barns from nonbarns. This assertion suggests that the relevance class associated with the process that one uses in the particular case should include all cases of barn recognition. On the other hand, it asserts that one uses the process only in certain situations and that one is aware of the difference between these situations and other cases of barn recognition. This second assertion suggests that the relevance class appropriate for evaluating the reliability of the process should include only the situations in which one uses that process. The moral here, I suggest, is that *task specifications, process characterizations, and relevance classes are everything in fairly evaluating reliabilism.*

4.) THE BONJOUR COUNTEREXAMPLES

As with many of the counterexamples to reliabilism, Bonjour takes himself to offer arguments against reliabilism in general, but considers only cases of reliabilism as a theory of justification. BonJour sets-out to confront "externalist radicals" (Armstrong, more specifically), his only weapon, his ability to illustrate

> the intuitive difficulty with externalism...[that] on the externalist view, a person may be ever so irrational and irresponsible in accepting a belief, when judged in light of his own subjective conception of the situation, and may still turn out to be epistemically justified,.... (BonJour 1980, p.59)

The two salient phrases in the above quotation are "intuitive difficulty" and "subjective conception." Indeed, though BonJour is thorough and conscientious in his writing, I think that

BonJour's examples are only convincing in the context of BonJour's rhetoric-driven intuition pump. The order and description of BonJour's examples play a large role in how compelling one finds BonJour's suggested intuitions. In fact, BonJour's examples all involve mischaracterizations or underdetermined characterizations of the task. Once one clears away the rhetoric and obscurants, BonJour's examples fail, I suggest, to illicit any strong intuitions not had by the reader previous to the examples.

All of BonJour's examples deal with cases of clairvoyance. Most adults exhibit a healthy scepticism concerning clairvoyance, making it a nice process to use if one wants to bias a reader's intuitions about the epistemic status of the process' outputs. I propose to consider a generic process, P, as I have no desire to bias readers against my position. Moreover, since I claim that order plays an important role in pumping intuitions, I consider BonJour's last example first: (BonJour 1980, p.62)

> NORM
>
> Norm generates an output, B_p, using the perfectly reliable process, P. Norm has no other data either for or against B_p in either its long-term or short-term memory.

Given the description of Norm's case, I take it to be contentious to claim either that Norm's tokening of B_p via P does or that it does not intuitively constitute his knowing that B_p. In fact, I think that BonJour agrees with me, as he immediately offers an argument that Norm does not know that B_p.[7] BonJour asks the reader to consider two distinct, but similar cases. In the first case Norm, actually has a prior belief that P is a reliable process, and this belief plays a role in the generation of B_p. So far, I would suggest, any intuitions one might have would tend towards attributing knowledge to Norm. But, BonJour interjects, suppose that Norm's belief about P lacks justification?

Since justification plays no role in my approach, I suggest altering BonJour's condition in a

way consistent with his own take on knowledge. BonJour holds that knowledge is justified true belief: Suppose, then, that Norm's belief about P does not count as knowledge. According to my account, in this alternative Norm case, Norm's tokening of B_p does not count as knowledge. That is, my approach requires that the represented data that Norm uses in generating his output count as knowledge. So, to generate a counterexample to my account, BonJour must now argue convincingly that in the case as it now stands, Norm's tokening of B_p counts as knowledge. Unfortunately for BonJour, his intuitions side with my approach in the Norm case as it now stands.

In the second case, Norm ostensively has no beliefs about P. This case looks like the original Norm case. BonJour introduces a false equivocation at this point: Norm's having no beliefs about P turns into a case where from Norm's "...standpoint, there is apparently no way in which he *could* know the President's whereabouts." (BonJour 1980, p.62) Somehow, though Norm has no beliefs concerning P, he cognitively registers that P could not be sufficient for him to know that B_p. I do not know how Norm could accomplish this feat without further reflection (i.e., using yet another process to evaluate P and/or B_p). Nor does BonJour offer any explanation of how Norm could have such a standpoint given that he has no beliefs about P or B_p.[8]

Barring further elaboration, then, I think it safe to assume that no one has any strong intuitions against my approach upon leaving the second alternative Norm case that they did not already have before considering the Norm case. I will now consider the examples that BonJour uses to prime his intuition pump: (BonJour 1980, p.61)

> MAUD
>
> Maud generates B_p using P, and B_p is veridical. P is reliable. Maud has a belief that P is reliable, and Maud's belief about P plays a role in her tokening B_p. However, Maud has no reason for her belief about P, and she maintains her belief about P despite massive counterevidence.

For BonJour, the key element of the Maud case lies in Maud's ignoring massive amounts of

"cogent scientific evidence" against her belief that P is reliable.[8] Interestingly, BonJour never mentions whether the process, call it $P_{m,}$ that gives rise to her belief about P is reliable. One needs only a minimal faith in science to claim that any process that would regularly generate outputs contradicting massive amounts of cogent scientific evidence is not likely to be reliable. After all, that is the whole point of BonJour's specifying that the evidence is massive, cogent and scientific. If he said that Maud had generated her counter evidence by flipping a coin, the sting of the evidence would slacken considerably. Again, as with the second alterative Norm case, Maud's tokening of B_p fails to count as knowledge on my approach, as her belief about P (which fails to count as knowledge) plays a role in her generating B_p.

At this point, I would also assert that BonJour bases his description of the Maud case upon a false supposition about the nature of belief and one's subjective perspective. BonJour wishes to claim that Maud believes in the reliability of process P despite what appears to her from her own subjective perspective to be massive, cogent, and scientific evidence against P. If one accepts the truism that one's belief in a proposition entails one's belief that it is true (hence the familiar example in philosophy of mind of a "belief box") and one has even minimal faith in human rationality, one must reject this example as psychologically impossible. That is, if one's believing a proposition involves one's cognitively registering its truth and one is aware of counter evidence one cannot subjectively view such counter evidence as showing the belief to be false. No one can plausibly claim "I believe the world is flat even though I clearly recognize that it is obviously not flat." Indeed, barring extreme psychological malfunction (even more extreme than schizophrenia), in such cases one invariably finds that the person has some reason to suppose that the counter evidence does not show the belief to be false. "I've seen it for myself," "someone I trust told me", etc. are examples of such views. Hence, this characterization of the case seems psychologically impossible.

Certainly BonJour does little to convince us of its psychological possibility for sane individuals.

I skip BonJour's Casper case, as it has the same essential features of Maud's case. I turn now to the case of Sam: (BonJour 1980, pp.59-60)

> SAM
>
> Sam generates B_p using P, and B_p is veridical. P is reliable. Sam has a belief that P is reliable, though he lacks evidence either for or against her belief about P. Moreover, he is aware of massive cogent evidence against B_p.

As I see it there are three versions of Sam's case consistent with BonJour's description. In the first version, Sam uses his belief about P to generate B_p. BonJour again omits the origins of Sam's belief about P. BonJour's omission makes it difficult to evaluate the effect of Sam's belief about P upon B_p. However, as in the Maud case, the epistemically dubious origins of Sam's belief about the reliability of P do cast doubts upon B_p's claim to knowledge.

In the second version, Sam uses P and only P to generate B_p. He then uses some other process, P_u, to evaluate B_p in light of the massive amount of cogent evidence against it.[10] Again, it seems uncontroversial that any process like P_u, that ignores massive amounts of cogent evidence against a proposition (having no supporting evidence) in evaluating that proposition, would prove unreliable. So, Sam's continued belief in B_p after the operation of P_u does not count as knowledge in the second version of Sam's case.

In the third version of Sam's case, BonJour might claim that Sam's process for belief revision, call it P_i, works in such a way that it *necessarily* accords the status of irrefutable or nearly irrefutable to beliefs generated by P. Such an additional stipulation, of course, undermines the intuition that BonJour wants to cultivate. If Sam's psychology is such that he *must* consider B_p irrefutable, it is difficult to accuse Sam of being "...irrational and irresponsible in accepting a belief, when judged in light of [her] own subjective conception of the situation,...." (BonJour 1980, p.59) Sam's

continued belief might seem irrational or irresponsible from our perspective. However, I would suggest that we would be hard pressed to maintain that Sam is irrational once we discover that P is a perfectly reliable process for evaluating beliefs and weighing evidence and P_i necessitates his continuing to hold that B_p. Why would we declare anyone irrational who must hold true beliefs to be true? Would we not consider Sam fortunate to have P and P_i? Would we not consider ourselves irrational to doubt Sam once we know the truth about P and P_i?

Let us consider a concrete case: Most people seem to find many of the verdicts of the probability calculus unintuitive. For instance, people often judge the conjunction of two events to be more probable than either event alone. The probability calculus, of course, instructs one that the probability of two events occurring together can be no more probable than the least probable of the events occurring in isolation. Now, we would be hard pressed to argue that someone who naturally reasons according to the probability calculus was irrational. They might seem to err in their judgements given our own intuitions. However, the appearance of error is due to a mistake on our part.

BonJour makes a similar move as the one just discussed in his second argument that Norm's tokening B_p in the original Norm case (where Norm generates B_p using only P) does not count as knowledge. BonJour asks the reader to consider a case where Norm also has a belief, B_a, for which he has large amounts of empirical evidence. Norm must choose between the beliefs based upon what he perceives to be their epistemic merit, knowing that a wrong choice would result in death. In this case, as with Sam, one can flush-out the case in two ways: In the first case, Norm uses P_u--the process that selects beliefs by ignoring massive amounts of cogent evidence--and chooses B_p. In such a case Norm looks very irrational. However, P_u looks very unreliable. So, BonJour's intuitions side with my account.

In the second case, Norm uses P_i, a process that necessarily selects B_p because P generated B_p. In this second case, Norm does not look irrational, as his psychology is such that he must consider B_p to have the greatest epistemic merit. Norm's choice might seem irrational or irresponsible from our perspective, but we would be hard pressed to maintain that Norm is irrational once we discover that P is perfectly reliable and P_i necessitates Norm's choice of B_p.

The final case I wish to consider is not mentioned by BonJour, but commonly offered by philosophers. Suppose that Norm drinks a potion that, unbeknownst to him, reorganizes his cognitive processes such that he can now visual see the number of objects in a given scene with absolute reliability even when there are huge numbers of these objects (this is sometimes called the rain man example). Upon quaffing the dregs of the cup he looks up at the bookcase and spontaneously forms the belief, "there are 200 books." He looks at the television and forms the belief, "there are 200,000 pixels." Surely, so the argument goes, Norm's beliefs do not count at knowledge, even though they resulted from a reliable process. After all, Norm himself would find these sudden beliefs mysterious, disturbing and without obvious epistemic merit.

Once again, the example confuses two processes: A visual process that generates the beliefs. This visual process is reliable. The process which sustains or undermines the beliefs in light of Norm's other beliefs (i.e., his beliefs that the beliefs are "mysterious, disturbing, and without obvious epistemic merit") is not the same process. If this second process is also reliable (as suggested by the example) and it undermines the belief (as is suggested by the very notion of belief), then the reliabilist is no more committed to calling the belief knowledge than the internalist. The internalist might object that the reliabilist must call the belief knowledge until the second process undermines the belief, while they (internalists) hold that it was never knowledge since Norm possessed defeaters to the belief even before forming it. However, I find this response, and all theories of justification

based exclusively upon evidential relations, unintuitive and inadequate for reasons dating back to Goldman (1986). The theory-laden nature of this response undermines it's force as a counterexample. To be effective, the irrationality objection must show reliabilism inadequate because the objection presents beliefs that are generated by a reliable process, but which appear--even to the believer--to be epistmically inadequate in a non-question-begging manner (i.e., based upon the believer's subjective perspective). The relations between beliefs in the above response are not part of the believer's subjective perspective.

5.) CONCLUSION

At this point, I have reviewed all versions of the irrationality objection of which I am aware. BonJour and the internalists have, I think, failed to make their case. The intuitions to which BonJour appeals in his book and article have, I claim, disappeared in my article. The reliabilist theory outlined in this paper can handle all of the cases that BonJour presents without making appeals to additional *ad hoc* or non-reliabilist clauses.

It seems safe, therefore, to locate the failure with BonJour and the interalists, and not with reliabilism. BonJour's most consistent failing lies in his characterization of the process(es) operant in the various cases and an implausible psychological assumption about the nature of belief. His discussion of the process(es) operant in the cases he describes falsely biases the case against the reliabilist.

In concluding I would offer an explanation for why these examples have been so robust and influential in philosophy. Internalists falsely suppose that since externalist theories like reliabilism tie the epistemic merit of a belief to properties other than one's subject perspective (or potential subjective perspective) of evidential relations between the belief and one's other beliefs one can easily separate one's subjective perspective from the claimed source of the belief's epistemic merit.

I claim that this is a false supposition for two reasons: First, the fact that one knows in virtue of believing places constraints upon the possible subjective perceptions of evidential relationships in ways which the counterexamples do not respect. One cannot believe something that one sees as obviously false. Second, subjective perceptions of evidential relationships are themselves a belief forming/belief sustaining process. These examples universally ignore this fact about perceptions of evidential relationships in order to create the illusion that the process that is stipulated to have perfect reliability is the process responsible for the belief's generation and/or continued existence in light of the subjective perception by the supposed believer of the damning evidential relationships.

BIBLIOGRAPHY

Armstrong, D. (1974). *Belief, Truth, and Knowledge*. Cambridge: Cambridge University Press.

Biederman, I. (1987). "Recognition by Components: A Theory of Human Image Understanding," in *Psychological Review*. 94:115-47.

BonJour, L. (1980). "Externalist Theories of Empirical Knowledge," in *Midwest Studies in Philosophy*.
5:53-73.

BonJour, L. (1985). *The structure of Empirical Knowledge*. Cambridge: Harvard University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Dretske, F. (1981). "The Pragmatic Dimension of Knowledge," in *Philosophical Studies*. 40:363-78.

Foley, R. (1985). "What's Wrong With Reliabilism?" in *The Monist.* 68:188-202.

Feldman, R. (1985). "Reliability and Justification," in *The Monist.* 68:159-174.

Ginet, C. (1985). "Contra Reliabilism," in *The Monist.* 68:174-187.

Goldman, A. (1976). "Discrimination and Perceptual Knowledge," in *The Journal of Philosophy*. 73:771-91.

Goldman, A. (1979). "What is Justified Belief?" in George Pappas ed., *Justification and Knowledge*. Dordrecht: D. Reidel.

Goldman, A. (1980). "The Internalist Conception of Justification," in *Midwest Studies in Philosophy*.
5:27-52.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge: Harvard University Press.

Goldman, A. (1988). "Strong and Weak Justification," in *Philosophical Perspectives*. 2:51-69.

Goldman, A (1991). “Epistemic Folkways and Scientific Epistemology,” in *Liaisons.* pp. 155-175. Bradford, MIT Press.

Goldman, Alan (1991). Unpublished comments on my paper at the 1991 Pacific Division *American Philosophical Association* meeting.

Haugeland, J. (1981). "Semantic Engines," in John Haugeland ed. *Mind Design*.

Cambridge: MIT press.

Marr, D. (1981). *Vision*. New York: Freeman and Company.

McGinn, C. (1984). "The Concept of Knowledge," in *Midwest Studies in Philosophy*. 9:529-34.

Pollock, J. (1984). "Reliability and Justified Belief," in *Canadian Journal of Philosophy*. 14:103-114.

Sanford, D. (1981). "Knowledge and Relevant Alternatives: Comments of Dretske," in *Philosophical Studies*. 40:379-388.

Schmitt, F. (1981). "Justification as Reliable Indication or Reliable Process?" in *Philosophical Studies*.
40:409-417.

Schmitt, F. (1983). "Knowledge, Justification, and Reliability," in *Synthese.* 55:209-229.

Sosa, E. (1985). "Knowledge and Intellectual Virtue," in *The Monist.* 68:227-244.

Sosa, E. (1988). "Beyond Scepticism, to the Best of our Knowledge," in *Mind*. 97:154-188.

Yourgrau, P. (1983). "Knowledge and Relevant Alternatives," *Synthese.* 55:175-190.

NOTES

1.) I only mention Goldman's theory in the text of this paper. However, Dretske's 1981 theory seems equally immune to BonJour's examples. Dretske's model for belief formation is that of a signal causing a belief. In order for a belief to count as knowledge, Dretske requires that the probability of the corresponding state of affairs given that type of signal *and what the system knows* equal 1. BonJour's most compelling cases require that the system know about massive amounts of cogent evidence against the belief or process, making it implausible that the probability of the state of affairs given the signal and what system knows will equal 1. Unfortunately, Dretske's approach has its own problems, which I comment upon in my dissertation.

2.) There are also independent worries about using *ex ante* justification to define knowledge. For instance, an imperfectly reliable set of rules, like a right set of J-rules, will permit some false beliefs. Do false beliefs counts as defeaters? Also, is *ex ante* justification recursive? If so, then false beliefs will justify more false *ex ante* beliefs, and so on. If *ex ante* justification is not recursive, why not?

3.) Feldman (1985) infleuntially points to the the characterization problem in criticizing reliabilism, though his discussion seems to conflate the characterization problem with the relevance class problem.

4.) (1) and (2) correspond to Marr's "Computational Theory" (1981, p.24-9). (3) corresponds to his "representation and Algorithm" and his "hardware Implementation" levels.

5.) The version of the red barn counterexample given here is based upon a version given by Alan Goldman in his comments on a paper I gave at the 1991 Pacific Division meeting of the *American Philosophical Association*.

6.) My account of Biederman's theory comes from his 1987 article. Biederman's theory does not include color primitives, though adding them does not seem to pose a problem for the theory. Also, Biederman's model assumes the solution to some problems of perception. For example, Biederman's model assumes a solution to the problem of distinguishing figures from background.

7.) BonJour has another argument to support his intuition in the Norm case. However, I postpone dealing with this example until the end of this section.

8.) Norm looks to be exploiting a second process to evaluate B_p in this case. I deal with such cases later in the paper.

9.) I should also note that BonJour's characterization of Maud's belief formation as involving other beliefs would seem to violate his characterization of the process as clairvoyance. Clairvoyance is neither an inference process nor a hypothesis evaluation process. It is the direct intuition of facts not present to the five senses.

10.) One can run a P_u version of Maud's belief about P. The verdict in such a case, however, again supports my approach.