

Revised Draft
June 2006

**Self Help for Epistemic Agents:
On two ways to rid yourself of pesky problematic beliefs ***

by Jayme Johnson
University of Massachusetts Amherst

§ 1. Introduction

As epistemic agents we have many beliefs. Some of these are beliefs about the world we inhabit, some are about abstracta, and some are beliefs about our beliefs. However, as epistemically fallible beings we have a number of beliefs that we have no good reason for having. Once this is realized, it will be natural for an agent to want to get rid of those beliefs which she has no good reason for having. She will want to improve her epistemic situation. Yet what does it mean to say of a belief that one has no good reason for having it? Before it can be made clear how to fix the problem it seems that one needs first to get clear what the proper diagnosis is. Does one have ‘no good reason’ for a belief only if that belief is false? Or should one’s having ‘good reason’ be a function the agent’s ability to come up with and give reasons for a belief? How one answers these questions will make a difference in how one thinks the project of epistemic self-improvement should be understood. Should it be the purpose of this project to increase the number of true beliefs within our belief set, or should it be to increase the number of justified beliefs?

In this paper, I examine the nature of the project of epistemic self-improvement as it is understood by a kind of internalism called Impatient Cartesianism (IC), and Hilary Kornblith’s claim that the IC understanding of such a project is critically flawed. I will closely examine two of Kornblith’s arguments against the plausibility of using an IC methodology to improve our epistemic situations, and how these relate to the IC understanding of justification. In doing so, an alternative, Kornblithian account of how the project should be understood will emerge; one that is integrally tied to his own conception of justification. I will then look at two closely related objections to Kornblith’s argument. One objection comes from Mark Kaplan, the other from Lawrence BonJour. Both objections suggest that there is an inconsistency between Kornblith’s theoretical obligations and his practice; that in arguing the way he does, Kornblith’s position seems self-undermining. I will then argue that while both of these objections touch on what seems to be a significant feature of Kornblith’s view, both suffer a problem of their own. Once

* I would like to thank Hilary Kornblith and James Ruvolo for the enormous amount of help they gave me on drafts of this paper. Most of what is good about this paper is due to their comments, criticisms and advice.

this problem is accounted for, I will attempt to formulate the worry as I see it, and in a way that avoids the shortcoming found in the Kaplan and Bonjour formulations. While exploring this dialectic, I will all the while try to relate it back to the project of epistemic self-improvement. What will emerge, I think, is a choice between kinds of epistemic advice, each with their own advantages and shortcomings.

§ 2. Justification, Truth, and Epistemic Self-Improvement

The questions I raised above suggest a connection between one's view on the relation between truth and justification and how one will, in turn, characterize what the project of epistemic self-improvement consists in. How one interprets what it means for an agent S to have or lack a good reason to believe P will be closely related both to how one will offer advice on how to maximize the number of beliefs an agent S has good reasons for, and how one criticizes competing advice.

2.1 The classic Cartesian Approach:

To set the stage for the dialectic of this paper, I will begin by looking at the classic Cartesian conception of justification, and how this relates to the Cartesian understanding of *the project of epistemic self-improvement* (hereafter the ESI project).

If one would have asked Descartes if the purpose of the ESI project was to maximize one's justified beliefs or true ones, he would have replied that there is no difference in these purposes. For Descartes, calling a belief "justified" requires absolute certainty in the truth of that belief. Thus, for Descartes, to increase the number of justified beliefs is just to increase the number of true ones. On this picture, in order for S to sufficiently dispel her doubt whether P, S must have Cartesian certainty whether P.

If one has *Cartesian certainty* that P is true then it cannot be the case that P is not true. The Cartesian notion of justification can then be explicated in terms of Cartesian certainty.

CJ: S's belief that P is justified if and only if S has Cartesian certainty that P. The purpose of the ESI project on this Cartesian understanding, then, is to uproot and discard all of our beliefs, and rebuild a new set of them, only allowing in those that are justified, and so guaranteed to be true. With that in mind, it can now be explained how the Cartesian ESI project is supposed to be carried out; it can be explained what Cartesian epistemic advice consists in.¹

¹ Given this manner of locution, providing an agent S with *epistemic advice* =df. supplying S with a mode or method in which to carry out the ESI project.

If S wants to follow Cartesian epistemic advice, then S begins by systematically doubting all of her beliefs. Through a process of introspection, S reflects on a belief P, and discovers what reasons (or lack thereof) she has for believing P. Fleshing this out, it seems that Cartesian epistemic advice has two components, a system to follow, and a mode of implementation for that system.

The system seems to be aptly characterized by the arguments-on-paper thesis (AOP).² Kornblith (1980) states AOP as follows.

Let's suppose that, for any person, it is possible, at least in principle, to list all the propositions that person believes. The arguments-on-paper thesis is just the view that a person has a justified belief that a particular proposition is true just in case that proposition appears on the list of propositions that person believes, and either it requires no argument, or a good argument can be given for it which takes as premises certain other propositions on the list (Kornblith 1980, p.599).

A more technical formulation might be given as follows:

AOP- S is justified in believing that P iff S believes that P, and there is some set of propositions $\{R_1, \dots, R_n\}$ such that S is justified in believing $\{R_1, \dots, R_n\}$, and there is a good argument from $\{R_1, \dots, R_n\}$ to P.

[We might simplify this by substituting the last conjunct with “ $\{R_1, \dots, R_n\}$ entails P”. For the purposes of this paper, no such substitution should be problematic.]

The mode of implementation for the Cartesian ESI project is reflection, or introspection. First, reflection is the mode by which one notices that one is in need of epistemic improvement in the first place. As Kornblith puts it, “nothing more than introspection is required to see that everything can be doubted” (Kornblith, 2002, p. 107). Additionally, reflection is used to reveal defects that have trickled into an unreflective agent's belief set, and to uproot and discard those defective beliefs and all that follow from them. Here the agent reflects on her belief that P, and scans her beliefs (as if they were on a list before her mind's eye). Once it is made clear that the reasons (the R's) for P are among S's beliefs, and these reasons are themselves justified, S is justified in believing that P.³ If there are insufficient reasons within one's set of beliefs, then P is

² This is not to suggest, of course, that Descartes would have put it that way. The philosophical era in which people put things in terms of AOP is off by a couple hundred years. In fact, while Descartes' view was a psychological one, and the AOP thesis, as I will discuss later, is taken largely as an apsychological one, it is striking how well AOP seems to illustrate the process that occurs during Cartesian meditations.

³ Leaving nothing to assumption, this process of reflective inference can then be further applied to these premise-like beliefs (the R's) which support P in order to make certain that these, too, are justified. Depending on whether one supports a foundationalist or coherentist view, this process will inevitably lead one to a set of foundational, axiomatic beliefs from which all inferential beliefs can be derived, or to a non-vicious circle of cohering beliefs. Note, however that on the classic Cartesian model under discussion

unjustified; S should decide to cease believing P. Hence, on the Cartesian understanding of the ESI project, “Introspection is thus called upon to reveal the defects of our current epistemic situation, to provide the will with the appropriate materials to remedy the situation, and to certify that those defects have been remedied” (Kornblith 2002, p.107-108).

There is one last thing to note about the Cartesian understanding of the ESI project. Note that the role of reflection is not directly justification providing. It is the *reasons* that an agent finds within her set of beliefs upon reflection which serve to justify a belief. The reasons justify; reflection is the method of access to those reasons.

2.2 Two reasons why most philosophers reject the Cartesian ESI project:

Without going too deeply into this issue, I would like to point out two serious problems for the Cartesian model described above. The first problem is that demanding Cartesian certainty for a belief to be justified seems to lead to total skepticism. It is widely held that the likelihood of ever obtaining this level of certainty about a belief is nearly impossible. Surely there is always room for *some* doubt whether a given belief is true. Second, even if one, in principle, could have Cartesian certainty about a given belief, it likewise seems impossible that anyone could put the Cartesian ESI project to work. The first step of following Cartesian epistemic advice is to rid oneself of *all* beliefs.⁴ It seems at the very least one would, in confirming that one has rid oneself of all beliefs, inadvertently but no less contradictorily form the belief that one has gotten rid of one’s beliefs. Snap. Hence it seems that no epistemic agent could actually follow Cartesian epistemic advice: it is impossible both to obtain the appropriate level of certainty to justify a given belief one is deciding whether to acquire, and it even if one could, it would be impossible to get the program started.

2.3 Two ways to reconcile the problems with the Cartesian ESI project

The inevitable result of giving up on Cartesian certainty as a criterion for justification means that any new theory of justification will be forced to recognize some gap between an agent’s true beliefs and her justified ones. A suitable theory of justification, then, must recognize that within an agent’s set of beliefs—regardless of whether an agent engages in *any* ESI project—there will likely be some justified true beliefs, some justified false beliefs, some true but unjustified beliefs, and some that are simply false and unjustified. Further, it will need to be

above, the process of giving reasons for belief stops in foundational beliefs, which Descartes described as “clear and distinct.” These particular theoretic details are unimportant to the current endeavor.

⁴ Or at the very least, one must *suspend* all belief. Either way, nearly the same point can be made.

shown how the inevitable presence of such a gap will affect the understanding of the ESI project. Again, it seems how one diagnoses the problem will affect how one thinks the project should be undergone. Moreover, as a result of giving up on entirely clearing one's belief slate as a starting point, any new theory of justification, if it is going to rely on reflection on one's beliefs for justification, will have to provide an explanation of where to begin the ESI project.

One attempt to reconcile these problems, while attempting to remain true to spirit of the Cartesian method, is undergone by those whom Kornblith has dubbed *Impatient Cartesians*.⁵ The Impatient Cartesian deviates from the classic Cartesian first by suggesting that we approach our beliefs with less initial doubt. The Impatient Cartesian realizes that many of our beliefs that were acquired unreflectively may actually be okay. We need only to pick out the faulty ones and correct those without reestablishing the ones that are safe. However, since the Impatient Cartesian also disagrees with the classic Cartesian about the level of overlap between justification and truth, something needs to be said about what will count as 'okay' and as 'faulty' belief from the perspective of IC.

While different Impatient Cartesians may have different views about what justification is, a strong point of agreement is that for a belief P to be justified, some kind of process of conscious reflection, or at least one's having the capacity to do so, is a necessary condition for one's knowing anything at all. Like the classic Cartesian, the Impatient Cartesian believes that S is justified in believing P only if S has good reasons for believing P. The reasons of interest here are ones found within S's own set of beliefs. The system, like before, is AOP, and the mode of implementation is reflection.

While we may get things wrong from time to time, the process of conscious reflection will, from the vantage point of the Impatient Cartesian, allow us to justify many of our beliefs that upon reflection, were found to be in need of justification. Such an exercise does not guarantee truth. No such exercise exists. But it does seem to point in the direction of true belief. Having justified beliefs, then, seems to get us closer to having true beliefs than we would otherwise be.

What counts as a 'faulty' belief from the perspective of the exponent of IC, I think, is one that is unjustified. S may never know for a fact that P is true, but if S is justified in believing that P, then perhaps that is enough to improve our epistemic situations. So the Impatient Cartesian understanding of the ESI project, if this is correct, is to maximize the number of justified beliefs

⁵ Kornblith notes that Chisholm, and at times BonJour are Impatient Cartesians. I will not here deal with the details of how this materializes in each of their views. For an account of this, see section 4.2 of Kornblith 2002. Suffice it to say that both suggest the venture of epistemic improvement i.e. the detection and removal of defectively held beliefs should be undertaken introspectively.

within one's belief set. This may not mean that all of these beliefs are going to be true, but then, we are still epistemically fallible, even if following an ESI program makes us epistemically better off. Hence it seems that much of the epistemic advice offered by the classic Cartesian is followed by the Impatient Cartesian. We might characterize the epistemic advice of the Impatient Cartesian as follows

ICA: For any belief P, If S is concerned about the justificatory status of P, then S should reflect on P, and if S's reasons meet the appropriate standards, then S is justified in believing P (and, of course, should keep believing P). If not, S should discard her belief that P.

Another way these two problems with the classic Cartesian understanding of the ESI project might be reconciled takes an externalist tack. Hilary Kornblith endorses such a view. The Kornblithian perspective agrees with the Cartesian idea that as strong a connection as possible should be maintained between truth and justification. Kornblith writes, "Like Descartes, I assume that the responsible epistemic agent is *truth-seeking* and that in attempting to improve his epistemic situation he is attempting to improve his reliability; he wishes to be more effective in *getting at the truth*" (2002, p.120).⁶ While Kornblith agrees that complete overlap between an agent's justified beliefs and her true ones is a pleasant fantasy, he is willing to change his notion of what counts as a reason for belief in order to maintain that connection between truth and justification. Like the Impatient Cartesian, on Kornblith's view there will be good reasons present when one is justified in believing P, but they may not be the same kind of reasons. To preserve the connection between truth and justification, Kornblith, like most externalists, spells out justification in terms of reliability. Such a view of justification can be put roughly as follows.

EJ: S is justified in believing that P iff P is produced by a reliable process.

Thus the Kornblithian understanding of the ESI project, as alluded to in the above quotation, is that the agent's goal should be to maximize the number of true beliefs. Since being justified in believing P in the externalist sense (P is reliable for S) increases the likelihood of P being true, then Kornblithian epistemic advice might be given as follows:

KEA: "If S desires that P be justified, then S should be responsive to situations in which P is reliably produced."

Unlike ICA, KEA does not include reflection as part of the advice, nor does it suggest a way to discard a belief lacking adequate reasons. Part of the reason for this is that Kornblith does

⁶ The emphasis is mine.

not think that S having further beliefs about the justificatory status of P automatically leads S to being further justified in holding P. Hence, the kinds of reasons needed to confer justification from the IC perspective don't seem to do much work. What makes a belief justified is its reliability, and that can obtain regardless of anything the agent thinks about it. This does not mean, however, that there are no reasons present here; it just means that what counts as a reason is different. Call these alternative kinds of reasons "Kornblithian reasons." I have no adequate technical definition to give, but there are, on BonJour's view, at least two conditions in which one might be said to have a Kornblithian reason for believing that P.⁷

KB1. If S believes that R and that P, and P is inferred from R by S, *and in fact the inference from R to P is reliable*, then R counts as a Kornblithian reason for S to believe that P.

KB2. If S believes that P and P is a non-inferential belief *reliably formed* as a result of some conscious experience R, then R counts as a Kornblithian reason for S to believe that P.

The proceeding discussion of Kornblith's views on the ESI project does not merely come as a positively given alternative to the IC understanding of the project. Much of it falls out of arguments Kornblith gives *against* the IC understanding of the ESI project. In the next section I examine two such arguments.

§ 3. Kornblith's critique of Impatient Cartesianism:

In this section I look at two arguments that Kornblith has raised against the Impatient Cartesian. The first argument rejects AOP as a way to understand justification. The second calls into question the reliability of using reflection for the purposes of the ESI project at all. It should first be noted that Kornblith's (1980) critique of the arguments on paper thesis was not targeting Impatient Cartesians specifically. It was meant to target the common element in most internalist theories of that time, coherentist and foundationalist alike.⁸ In studying Kornblith's work, I judged that AOP was an important element in the Impatient Cartesian's understanding of the ESI project, and wove this earlier critique into the one found in Kornblith (2002) (which I discuss in 3.2.) If in what follows I lapse into speech which suggests that Kornblith had it in mind to treat

⁷ See BonJour "Kornblith on Knowledge and Epistemology" *Philosophical Studies* (2006) 127: esp. pp. 331-332.

⁸ Putting the point in this way is not quite right either, since internalist/externalist distinction had still not quite made its way into the literature at that time. However, the distinction does seem accurately to pick out internalism as Kornblith's target in that paper.

these two arguments as a joint attack on Impatient Cartesianism all along, this should be taken as an artifact of my writing, and not of Kornblith's actual intent.

3.1 The arguments-on-paper thesis

As I have framed the dialectic, recall, AOP is the system used to scrutinize one's beliefs in both the classic Cartesian and the Impatient Cartesian understanding of the ESI project. I formulated AOP above as follows.

AOP- S is justified in believing that P iff S believes that P, and there is some set of propositions $\{R_1, \dots, R_n\}$ such that S is justified in believing $\{R_1, \dots, R_n\}$, and there is a good argument from $\{R_1, \dots, R_n\}$ to P.

AOP reduces the question of epistemic justification to a question of what counts as a good argument.⁹ In this sense, the thesis is apsychological; whether or not an argument is a good one does not rest on anything about the agent other than what beliefs are on his list, and perhaps the propositional contents of those beliefs. Kornblith's argument against AOP runs as follows:¹⁰

A1. Consider Alfred. Alfred justifiably believes that p, and justifiably believes that if p then q;

(1). If AOP is true, then given A1, necessarily, Alfred is justified in believing that q.

On any reasonable theory which endorses AOP, (1) will be true given the assumptions underlying Alfred's case. Both the belief that p and the belief that if p then q are on Alfred's list; they are members of his set of R's. Further, the inference from these to q would make any reasonable list of allowable inferences. "It is clear, however, that Alfred need not be justified in believing q. Alfred may well be aware that he believes that p and he believes that if p then q, and yet fail to believe q *on these grounds*" (602).¹¹ Thus,

(2) Its not the case that, given A1, necessarily, Alfred is justified in believing that q.

⁹ And what counts as a good argument will vary depending on the theory in question. Foundationalist theories will want the R's to be foundational beliefs, or at least be beliefs that are themselves capable of being derived from foundations by good arguments; thus in addition to a list of propositions, a further list of acceptable logical inferences is needed. For coherence theories, what counts as a good argument will be function of how the belief can, perhaps by some logical principles, be shown to stand in the coherence relation to the R's.

¹⁰ For Kornblith's actual formulation see his (1980) pp. 601-602.

¹¹ The emphasis is mine.

From this, Kornblith concludes, “Alfred is justified in believing that q only if his belief that q *depends* on his beliefs that p and if p then q. The notion of belief dependence cannot be accounted for in terms of the contents of the various beliefs held. The arguments-on-paper thesis is thus false” (602). The problem here for AOP conferring justification on Alfred is that AOP fails to take into consideration the causal basis of Alfred’s belief that q. Alfred might have the appropriate beliefs on his list that *would* justify his belief that q, but it should not be claimed that Alfred is justified in his belief that q unless it is, *in fact*, those reasons found among his set of R’s that causally sustains his belief that q.

An agent following ICA will reflect on her beliefs in order to judge whether she has adequate reasons. In doing so she will be implementing AOP to confer justification on beliefs she believes are in need of justifying, beliefs she may have formed when she was being unreflective. If Kornblith’s critique of AOP is correct, then there seems to be beliefs that will pass through this process and get counted as justified, but perhaps should not be. If so, then ICA begins to sound like bad advice. In fact, the problem may even be worse than this. More than just implementing an unsatisfactory system in which to scrutinize our beliefs, it may actually be the mode of implementation itself that is problematic. Kornblith’s second objection to the IC understanding of the ESI project calls into question an agent’s ability to effectively use reflection as a mode of getting at her reasons for belief.

3.2 Reflection is unreliable at detecting errors in our beliefs

The clearest statement of Kornblith’s argument can be found in chapter 4 of his *Knowledge and Its Place in Nature*. To understand his argument, it must first be recalled that on Kornblith’s understanding of the ESI project, the agent seeking to improve her epistemic situation is interested in maximizing her justified beliefs only to the extent that this justification will make her beliefs more likely to be true. Hence, for Kornblith, agents seeking to improve their epistemic situations are seeking to have more *true* beliefs. The problem with ICA, claims Kornblith, is that the kind of justification one gets from it does not make truth more likely. In fact, Kornblith thinks, following ICA can make an agent believe that her epistemic status has improved, but that this is a false sense of security. The problem, like the problem pointed out for AOP, is that reflection often fails to supply us with our actual grounds, and therefore fails to reveal to an agent what her actual reasons for belief are.

Kornblith begins the argument by consulting data from a number of different psychological experiments. All of these data are then analyzed in such a way so as to support two claims.

(K1) Once an agent has formed a problematic belief, it is terribly difficult for the agent to detect introspectively that belief as problematic [and therefore to appropriately uproot and discard that belief].¹²

Given Kornblith's understanding of the ESI project, calling a belief that P *problematic* in this context describes cases in which P is formed either as the result of a faulty inference, or because of a perceptual error. Perhaps another way to put it would be to say that S's belief that P is *problematic* iff P is unreliably produced in S.

(K2) Even if the reasons an agent gives for holding a belief are good ones (and they typically are, or at least would be), in cases in which the belief is problematic, these reasons are not typically the ones for which the agent holds the belief.

While an agent will often rationally reconstruct what she takes to be good reasons for having a problematic belief, in many important cases, these will not be the *actual* reasons the agent has for holding that belief. The reasons would be good ones if they were actually hers. The problem is that they often are not. (Indeed, if the reasons the agent gave in these kinds of situations actually *were* reasons, then the belief probably would be justified.)

From these two claims, it follows that when put to the task of epistemic self-improvement—a task by and large comprised of engaging in a process of isolating and rationally scrutinizing suspect beliefs—our introspective/reflective capacities come up quite short. Without the confidence that the reasons employed to justify the belief will be one's genuine reasons, and without any good sense of which beliefs even to hold suspect, reflection cannot be reliably used to make corrections in one's set of beliefs. Since reflection is unreliable as a tool for detecting errors among one's set of beliefs, it is therefore incapable of being used effectively to improve one's epistemic situation.

Now the Impatient Cartesian will want to reject both K1 and K2. First, in order to pick out a faulty belief, an agent must first recognize that belief as problematic. For such an endeavor to be successful, the Impatient Cartesian must reject K1. For ICA to be effective, it must not be the case that it is difficult to use reflection to identify suspect beliefs. Acceptance of K1 means the Impatient Cartesian loses the short-cut that makes the view an improvement on the original Cartesian prescription of scrutinizing *all* of our beliefs. If one cannot put a finger on which

¹² Note that this claim does not rule out cases in which an agent can reliably recover her reasons for belief. If S's first order beliefs are, for the most part, accurate, while S will have a hard time uncovering unreliably formed beliefs, S will still be able to reliably cite reasons for perhaps many of her reliably produced beliefs. The problem, one might say, is that the bad beliefs look as good as the good ones; but this is not to say that there are no good ones.

beliefs are suspect in the first place, one cannot afford to skip a single one. Second, the Impatient Cartesian will also want to reject K2. The act trying to justify our beliefs reflectively with the reasons we have for believing them requires us to have reliable reflective access to our actual grounds for belief. In short, the Impatient Cartesian needs it to be the case that when we reflect on our reasons, we can accurately pick out which ones we really have. If an agent cites reasons for belief that are not her own she is not justifying that belief at all, she is rationalizing it. Since an improved epistemic situation seems to be one in which an agent has minimal rationalized beliefs, if K2 is true, then ICA will not work. The conjunction of K1 and K2, then, seems to spell disaster for the Impatient Cartesian's ESI program. In short, if K1 and K2 are true, then ICA is bad epistemic advice.

Kornblith supports K1 and K2 not with *a priori* principles, but with tomes of empirical data from cognitive psychological studies. Since the purpose of this paper is not to call into question the way in which Kornblith gets from the data to the premises of his argument, but is rather to place this argument within the current dialectic, I will leave much of the particular details of the experiments aside, and focus on how Kornblith draws the premises from general points about the data.¹³ The following passage seems to summarize nicely what Kornblith takes to be the salient points from the experiments.

Subjects who undertake the project of epistemic self-evaluation by attempting to introspect the source of their beliefs will, at least in the cases described, misdiagnose their reasons for belief.[...] Processes that are in fact unreliable are not recognized as such. Factors that should not influence a subject's judgment are not recognized as influential. In cases such as these, introspection is powerless to detect the errors we make, but in misdiagnosing the source of our judgments and our reasons for believing, the reliance on introspection as a tool for self-evaluation merely instills a false sense of confidence in an already misguided agent (*Ibid.* p.115).

To show how these data serve to establish the truth of K1 and K2, Kornblith makes the following two points. First, the data strongly suggest that "our processes of belief acquisition, and indeed mental processes in general, are largely unavailable to introspection" (*ibid* p.120). If an agent cannot introspectively access the sources of her beliefs then she will have a very difficult time using reflection to determine which of her beliefs were formed by unreliable processes. From this, it follows that K1.¹⁴

¹³ This is not to suggest that there is not an interesting project here. It would be quite worthwhile to systematically look at the conclusions Kornblith draws from these studies, explore the literature for alternate interpretations, and compare the studies Kornblith considers to studies that seem to come up with contrary conclusions.

¹⁴ An interesting side note: Adam Elga, in his "On Overrating Oneself...and Knowing it" has suggested that some of the beliefs an agent has about herself she will not discard even when she finds out they are

Second, introspective reports require rational reconstruction, and the data suggest that this, in turn, involves “heavy reliance on the presupposition of our own reliability” (120-121). When an agent is asked why she holds a particular belief, she will typically begin by assuming that she is rational, and then give reasons that seem to her to be rational support for that belief. For example, in the Nisbett and Wilson (1977) cases the agents gave reasons that a rational agent would give for preferring one product over the other; the features of the garment the subjects would cite as reasons, of course, were not really the agent’s reasons. The real reason had to do with the position of the product, not features of it. The only explanation for why the subjects gave the reasons they did is that, while they may not have been aware of doing so, they must have presupposed their own rationality and constructed the reasons. Since such rational reconstruction is so integrally tied to reason-giving, it follows that K2.

Given that K1 and K2 seem so hard to deny, at this point the only tools available to the Impatient Cartesian to dispute Kornblith’s critique are his own intuitions about the connection between reliability and justification. The Impatient Cartesian may ask, “Why suppose there is any sort of necessary connection between reliably produced beliefs and justified beliefs? All that has been shown above is that following internalist advice will not improve reliability, not that those beliefs aren’t justified.” Kornblith’s reply to such an objection is that one is entitled to separate reliability from justification, but “if one does so divorce one’s concept of justification from that of reliability, it becomes very hard to say why one should care to have justified beliefs” (*ibid.* 119). If the process of epistemic improvement is to take place on the grounds of producing more justified beliefs and getting rid of those beliefs that lack proper justification, and justification does not, at the very least, make truth likely, then it is hard to see why this endeavor would do anything to improve one’s epistemic situation. Reliably produced beliefs are ones that are very likely to be true. If one’s conception of justification does not offer that if a belief has it, it is likely to be true, then calling a belief “justified” on such an account might be well and good, but it will not serve well to improve our epistemic situation.

Hence, for an Impatient Cartesian to continue to maintain that the truth of K1 and K2 does not lead to the conclusion that ICA is bad epistemic advice, he must maintain that there is no connection between reliability and justification, that there is a worthwhile conception of justification that is completely divorced from truth, and that having such justification has a value in improving our epistemic situations.¹⁵ In short, the Impatient Cartesian must explain how it

likely to be false. The reason seems to be that it is psychologically healthy to have an inflated sense of one’s own abilities.

¹⁵ Some internalists deny that the connection between truth and justification is severed in the way suggested here. While the connection may not be the straightforward one reliabilists insist on, these philosophers

makes for good advice to prescribe an agent with a procedure that will result in that agent having a false sense of epistemic security in the quality of her beliefs.

§ 4. Two Objections to Kornblith's line of argument

In this section I look at two objections. The first objection, from Mark Kaplan, is raised against Kornblith's critique of AOP. The second objection, from Lawrence Bonjour, is raised against Kornblith's argument that reflection is unreliable at detecting errors among our beliefs. Both objections center on the idea that in arguing in the way he does, Kornblith's position seems self-undermining.

4.1 Kaplan's objection

Kaplan begins his critique by confessing of three beliefs, out of many he has like them, that he believes they are justified even though he has no idea what their causal origins are, or what causally sustains them in him (see p. 352).¹⁶ The following scenario describes such conditions:

S believes that P; S believes that (S is justified in believing P); but it's not the case that S knows whether her belief that P has a "happy etiology."

Kaplan says that despite the fact that he does not know the etiologies of his beliefs that the above scenario describes, his intuitions tell him that he is justified in believing them anyway, or at least that he ought not to feel as though he should rid himself of these beliefs solely on the grounds that he does not know their etiologies. Kaplan can think of reasons why he believes these beliefs are justified, despite his inability to recall the origins of these beliefs. Hence, the intuition at work in Kaplan might be put as follows.

INT1. If S is justified in believing that P then S can produce upon reflection, her reasons for believing P.

However, suppose that one's theory of justification is centered on "specifying the features a process must have in order for the beliefs it causes and/or sustains to count as justified and to identify processes which have this feature" (Kaplan 351). If one is going to hold a theory of justification like the kind just described, then given INT1, it seems to follow that in order for a belief to be justified on such a theory, the agent would need to know, or could remember if she thought about it, the etiology of her belief. Advocates of the kinds of theories of justification just described—reliabilists—however claim that no such information is needed to secure justification:

maintain, there still is some meaningful connection between the two. See, for example Feldman, Richard. "Chisholm's Internalism and Its Consequences," *Metaphilosophy* Vol. 34, No.5, (October 2003) 603-620.

¹⁶ Kaplan, Mark. "Epistemology Denatured," *Midwest Studies in Philosophy*, XIX (1994) 350-365.

a belief P is justified just in case P is produced by a reliable process, and knowing things about that process need not make affect the justificatory status of P.

As a result of this claim, Kaplan thinks that naturalists [reliabilists] are committed to something like the following conditional:¹⁷

If “S is justified in believing P” means nothing more than “S formed P by a reliable process” [and it is not required of S that S be able to explain the causal origin of her belief that P for P to be justified], then S will be unable to tell which beliefs S ought to divest herself of, and which beliefs she ought to maintain.

Naturalists, Kaplan thinks, are committed to this conditional precisely because their theories produce the result that an agent will not be able to distinguish which beliefs have “happy etiologies” from the ones that do not. He writes, “For, insofar as I find myself in no position to say, of any of the three beliefs to which I have confessed, whether it enjoys the fortunate etiology required for it to count as justified by natural lights, then I cannot but conclude that, should I embrace the naturalist credo, I will find myself in no position to say, of any of these beliefs, whether that belief is justified (*ibid* 352).” If so, Kaplan continues, then under naturalism beliefs whose causal origins lie in today’s horoscope look just as good as ones tracing to reliable sources. Kaplan thinks there is something entirely unsatisfactory about this. The problem is that reliablism does not preserve INT1. In short, if justification is reliability, then in many cases the agent will not be able to reflectively access her reasons for P (namely, S will not be able to recall P’s etiology); so there will be cases, from the reliablism perspective, in which the antecedent of INT is satisfied, but the consequent is false.

Kaplan claims that it is natural for agents to want to rid themselves of beliefs for which they have no justification, and on the naturalist account, an agent finds herself in no position to determine which of her beliefs to keep, and which to discard. Not feeling confident that a belief is justified seems good enough of a reason, to Kaplan’s mind, to want to divest oneself of that belief. “But this is methodological madness. It is madness to suppose that I should be trying to divest myself of these beliefs. And that is because it is madness to suppose that I *need* to be in a position to say that a belief of mine enjoys a happy etiology before I can feel that it is intellectually respectable to harbor the belief” (*ibid*. p. 352).

Kaplan suggests that the naturalist would not be worried by this. The reason why a naturalist would not be moved by this, he thinks, is that the naturalist will claim that his theory of

¹⁷ The point is really pinned to naturalists to the extent that naturalists are externalists, or more specifically, externalist reliablism.

justification is nonregulative.¹⁸ Call a system *T* *regulative* just in case *T* provides a set of rules for an agent to follow in the conduct of inquiry, and *nonregulative* just in case it is not regulative. Since the naturalist merely offers a theory of what a justified belief *is*, claims Kaplan, and not a procedure for an agent to follow in order to engage in the activity of justifying her beliefs, it doesn't really matter with respect to the justificatory status of a belief whether *S* can tell whether it is appropriately grounded or not. Such concerns only matter if a theory of justification is regulative.

However, by making one's theory of justification nonregulative, Kaplan suggests, one commits oneself to a view that gives the term "justified" no methodological import, "—that there is nothing in the set of rules to be consciously adopted by an inquirer that calls for her to determine whether any of her beliefs is justified in the nonregulative use of that term" (353). And if it is characteristic of the nonregulative use of the term "justified" that it has no methodological import, Kaplan further claims, then

- (1) "it is hard to see what point there could be to a system of evaluation dedicated to saying when it is that a person's belief is justified in the nonregulative sense"
- and
- (2) "it is hard to see on what basis one could judge (as the naturalists do in their critique of the arguments-on-paper thesis) that the expression, "justified" is being incorrectly applied" (*ibid.* 353).

Kaplan then suggests in support of (1) that commitment to any nonregulative theory of justification, no matter how draconian, is commitment to nothing more than a style of epistemic appraisal, a style "whose judgments are of absolutely no import to our conduct of inquiry" (*ibid.* p.354). Without any sort of import to our conduct of inquiry, even if modes of justification that seem to accord with our intuitions fail on the theory, it matters little, and seems further to matter little how we actually conduct our own business of belief justification.

With no facts about what we are prepared (and not prepared) to believe capable of undermining any of our proposals, we can endorse any one we care to (or any pair or trio—why not commit ourselves to a variety of nonregulative styles?) without fear of having made a mistake. A mistake is only possible when there is some constraint on what counts as getting matters right. [...] Cut loose from the obligation to answer to our practice of inquiry and criticism, proposals about how to make nonregulative judgments about what beliefs are justified answer to nothing (354).

I think Kaplan's first point is that, if no advice can be given to an agent on how to determine whether her own beliefs are justified, then it will be pointless for her to engage in any

¹⁸ Kaplan notes that the distinction between regulative and nonregulative evaluative systems comes from Goldman. He points the reader to Goldman's *Epistemology and Cognition* (25-26) and also his "The Internalist Conception of Justification," (28-29).

activity of inquiry and criticism of beliefs. (If S is can't tell which beliefs are justified, why should she care about justification?) What Kaplan seems to be suggesting is that if non-regulative theories offer nothing to the first person perspective in epistemic inquiry, then they are useless. Relating this back to the dialectic of the paper, if it is a feature of a theory of justification that an agent cannot use it to tell which of her beliefs are justified, then the agent cannot engage in any ESI project that is based on that theory of justification. In short, naturalist accounts of justification, in that they remove the agent from any direct role in the activity of belief justification, remove any hope of an agent following epistemic advice of any kind.

In support of (2) Kaplan submits that just this kind of move that the naturalist makes can save the arguments-on-paper thesis from the criticisms that have been put forth against it. In Kornblith's treatment of AOP, what does the work to disparage AOP's ability to confer justification on an agent is an ordinary intuition. The intuition Kornblith draws attention to is that there is no reason to suppose that Alfred necessarily believes *q* on the basis of the contents of his other beliefs.

But now, equipped with the regulative/nonregulative distinction, it is easy to say why the intuition appears to have such force and why the account it would seem to undermine nonetheless manages to emerge unharmed. The apparent force of the intuition, we can say derives from the fact that, taking "justified" in the regulative sense, the intuition is quite correct: no sound methodology worthy of the name would sanction forming a belief in the manner described. But, we can say, the arguments-on-paper thesis does not claim otherwise. The arguments-on-paper thesis is nonregulative in nature: it is committed only to the claim that in the nonregulative sense of "justified", her belief is justified. And this claim implies nothing about what sound methodology sanctions. [...] Theses about justification in the regulative sense are fated to survive only at the pleasure of our methodological intuitions. But theses about justification in the nonregulative sense are answerable to none." (356)

Kaplan seems to be saying that in order to save his theory from the charge of methodological madness, the externalist commits himself to a view that completely removes the force from his objections to AOP. By having a theory of justification that denies INT1, which on Kaplan's view seems to be any and all nonregulative ones, an agent has no means by which to evaluate the argument given against AOP. In fact, thinks Kaplan, an agent cannot very well evaluate any argument against any theory so long as the agent is taking "justified" in the nonregulative sense. The act of evaluating an argument, Kaplan seems to be suggesting, requires a first-person role of reflecting on the argument's premises, and the inference from those premises to the conclusion. In other words, in order to be justified believing the conclusion on the basis of the premises, "justified" must be taken in the regulative sense; INT1 must be preserved. Kaplan thus concludes that "maintaining that the naturalist account of justification is

nonregulative saves it from the charge that [it] is methodologically mad only at the cost of rendering the account empty” (356-357).¹⁹

Kaplan then raises the further objection that naturalists run directly into this same problem of inevaluability when giving arguments in favor of their own position on justification over any sort of internalist or aprioristic approaches. With regard to the naturalist’s positive position, Kaplan remarks that he is quite sympathetic with the claim that “the fact that a methodology accords with our intuitions concerning rationality and evidential support is entirely compatible with its being quite unreliable at delivering us cognitively virtuous states in our world...” (359). Yet he is struck by the fact, that in support of such a position, the naturalist offers a series of arguments, as opposed to some scientific evidence that “were we to believe what they would have us believe as a causal consequence of their having exposed us to these arguments, our belief would be the result of our having instantiated a process that is apt to have a high frequency of generating epistemically virtuous states in human beings in our world” (359-360). But why accept these arguments? Why think they are cogent? On what grounds ought these arguments to be evaluated? Kaplan answers that “It is hard to see what we can do except evaluate these arguments by the light of the very sorts of epistemic intuitions which the naturalists are so eager to disparage” (360). By not preserving INT1, the externalist not only undermines his ability to give arguments against AOP, he undermines his audience’s ability to evaluate any arguments he might give in favor of his view of own view of justification as well. If one takes “justified” to be a nonregulative term, then in giving arguments in favor of one’s own view, like in the case above, there seems to be no clear way in which to evaluate those arguments. The audience has been cut off from the very sorts of intuitions needed to evaluate the argument.

Kaplan’s argument might be summarized as follows: There is clear link between INT1 and “justification” in the regulative sense. One’s theory cannot both disparage INT1 and include a procedure providing a means of first-person regulatory access to one’s beliefs (be regulative). To try to have both is methodological madness. To reconcile, one must either embrace INT1, or deny that one’s theory is regulative. Embracing INT1 is denying that justification is reliability. Denying that one’s theory is regulative is to take oneself out of the race, so to speak. Nonregulative theories deny the role of reflective reason giving for justification, so an exponent of such a theory cannot provide reasons to prefer her theory over another, or justify arguments

¹⁹ Note that Kaplan does not take this to be a defense of AOP. He claims himself that he has no brief for AOP (p. 351) Thus at this point in the paper, Kaplan reminds the reader that this is not to say that the AOPers do any better. They suffer the same methodological dilemma as the naturalists do. The naturalists have it right that AOP is problematic, but they have it for the wrong reason. The problem with both AOP and the naturalist critique of it, if Kaplan is right, is that neither is committed to a view of justification that is action-guiding.

against competing theories. Thus, for Kornblith to disparage INT1 and give arguments both against AOP and in favor of his own view seems self-undermining. By denying the role of INT1 in justification, he loses his ability to provide good reasons for accepting his arguments.

4.2 An initial reply to Kaplan:

As I see it, the first part of Kaplan's objection boils down to the claim that for a theory of justification to be worth anything, it must be regulative. This becomes even clearer when one considers passages like the following. "Properly conceived, the job of a theory of justification is just to provide rules for (and insights into) the proper conduct of inquiry" (Kaplan 358). Without the ability to provide rules and guidelines in order to steer an agent toward making epistemically responsible decisions (in terms of what beliefs to accept or reject,) that theory is methodologically inert, and can make no argumentative claim to be better than any other nonregulative theory. So the dilemma that Kornblith and others seem to be put in here is that they must either make methodologically mad claims, or go nonregulative about justification and have empty, inevaluable theories. However, I think that Kaplan errs in framing the dilemma in this way. In this section I will address the first part of Kaplan's critique.

First, is it really methodological madness to suppose that we have beliefs the "happy etiologies" of which we no longer (or never did) have introspective access to, but that we are still justified in believing? Kaplan's supposing that it is madness seems to rest in the following assumption: being justified in believing P means having the confidence to say that we are justified in believing P. Having confidence that we are justified seems to be a product of being able to give our reasons for belief. This suggests that Kaplan thinks that for a theory of justification to be proper, it must preserve INT1. But this just sounds like an argument saying nothing more than "my intuitions tell me that naturalistic accounts of justification are wrong." Support for the intuition seems to come from the claim that externalists cannot give reasons at all. But just isn't right. The externalist merely maintains that our ability to give introspective reasons (the kind needed for the inference that Kaplan suggests is indicative of justification) for a having formed a belief p is neither necessary nor sufficient for that belief to be justified. This does not mean that an agent cannot be justified in believing that p and still provide such reasons. It merely means that she doesn't have to, and sometimes when she does, they don't justify it. The perspective that emerges is that there will be many beliefs we have that we, in fact, will be justified in believing, despite our inability to give internalist reasons. To insist, then, as Kaplan does, that it is madness if a theory does not preserve INT1 seems already to presuppose that externalism is false, or at least methodologically mad. To my mind, there is nothing

methodologically mad about supposing that I have beliefs which are justified, but that I do not have any good internalist reasons for believing.

Second, the move Kaplan makes from claiming that all naturalist theories, in denying INT1, are nonregulative, to the claim that because of this are all completely methodologically inert, is much to fast. Kaplan's assumption here is that a theory of justification either provides a decision procedure featuring a first-person privileged role in evaluation, or it is completely pointless. To my mind this is just a false dilemma. It is true that naturalists deny INT1, and as a result, deny a first-person privileged role in justificatory evaluation, but this does not *ipso facto* make such theories methodologically inert. There might be a perfectly clear third-person role in justificatory evaluation by which naturalist theories are quite methodologically dynamic. Kaplan fails even to entertain this possibility. Without a clear argument against the possibility of using nonregulative theories of justification for third person judgments, it seems that one need not grab either of the horns Kaplan presents.

The further question should also be raised whether all externalists are committed to a nonregulative view of justification as Kaplan alleges. It seems that the regulative/nonregulative distinction as Kaplan frames it divides theories of justification into those that demand internalist reasons for justification and those that do not. If all Goldman meant by this distinction in the first place is that externalist theories of justification deny the role of internalist-brand reasons in conferring justification on an agent, then Kaplan's demand that for a theory of justification to be worth anything it must be regulative is just to beg the question in favor of internalist theories of justification. Perhaps there is a sense of "regulative" such that externalist theories are regulative. If "regulative" is taken to mean "able to provide some kind of epistemic advice to the agent" then Kornblith's theory seems to make the cut. While the advice offered is not ICA, there is still advice given. One might characterize that advice as "when you form a belief, if you want it to be justified, be responsive to situations in which that belief is reliably produced."²⁰ Does this advice guarantee that there will be internalist reasons available such that the agent can call upon them? No, but to claim that there must be is already to suppose that the theory is false.

Thus, there seems to be problem with the first part of Kaplan's critique. It remains to be seen how this will affect the second objection, namely that without providing the kinds of reasons that Kornblith rejects as justification-giving, his argument is inevaluable. Since this part of the critique is so similar to Bonjour's I will first take a look at his objection, and then treat the two together.

²⁰ Note the similarity here to KEA above.

4.3 BonJour's objection:

BonJour begins his attack by raising two questions about the studies upon which Kornblith bases his arguments against the Impatient Cartesian, and which serve as evidence for his positive position.

First, do the investigators in question have any good *reasons*, reasons that they can be reflectively aware of and could in principle cite to others, for believing that the alleged results of the study are in fact probably correct? Second, do those same investigators have any good *reasons* for believing that the methods that were employed in the study are in fact reliable? Does Kornblith himself have any good *reasons* for believing either of these things? (329).

Next, he points out that while it is entirely possible that the results of the studies and the claim of reliability may give externalist brand justification, and even (externalist) knowledge, “this is all quite compatible with no one, neither the investigators, nor Kornblith, nor anyone else having any good *reason* for believing this...” (BonJour, 2006, p.330).

He then makes two points about what it would take to have a good reason to believe these things. The first point is that, in giving such reasons, one must be responsive to the fact that the regress problem applies as much here as anywhere. Determining whether or not such reasons can be given without falling into the regress problem, BonJour thinks, cannot be done without a retreat to traditional epistemology. The reason BonJour thinks this is that for any belief, if one gives reasons for that belief—ones of which the agent is reflectively aware—then the reasons for those reasons must themselves be called into question. Thus, in order to avoid the regress of reasons, this exercise must terminate into adequate foundations, or must close in on itself in a non-vicious circle of coherent beliefs. For BonJour, currently, this means determining adequate foundations.²¹ Determining whether this has occurred, or can occur, is largely, an a priori, reflective practice. Even if we suppose that foundationalism is correct, and reasons are available for the studies and for Kornblith's claim, none of these are foundational. Securing foundations, again, is an exercise that must be performed reflectively, by drawing inferences. This leads to BonJour's second point; it needs to be shown that the inferences from the beliefs under discussion to adequate foundations are themselves cogent. “The traditional view, which I believe to be correct, is that the reasons that support the cogency of the inferences if there are any, must be *a priori* in character.” (330-331)

²¹ Since Kornblith and BonJour agree that coherence notions of justification are problematic, I omit discussion of what would count as a good reason from the coherentist perspective.

Kornblith, as Bonjour rightly notes, rejects that a priori considerations are needed for justification; reliability, Kornblith will say, is all that is needed. Bonjour suggests, however, that reliability is not reason providing. He claims that even if the inferences are reliable and the premises are true, there still might be no reason to think the conclusion is true. Now Bonjour, acknowledges that there might be Kornblithian reasons available (as discussed in § 2 above), but he thinks that the presence of such reasons will not meet the demands asked of them. What Bonjour says specifically about a Kornblithian reason is that “unlike an internalist reason, [it is] not in general open in any useful way to critical assessment by others to whom it is offered as a reason nor by the person offering it himself” (Bonjour 2006, p.332). Since Bonjour himself states very clearly what he takes the problem to be for a view that rejects the need for internalist reasons, I quote him at length.

Suppose that the author of one of Kornblith’s studies puts forth its conclusion and when asked for his reasons, offers various other things that he believes. If the inference from those other beliefs to the conclusion in question is reliable, the reason is a good one; if the inference is unreliable, it is not. But if neither the investigator nor the audience have any reason access to the fact of reliability or unreliability, it is impossible for any of them to tell which of these possibilities is actually realized. Of course, they *may* have further beliefs about the reliability of this first inference, but the same problem recurs for those beliefs as well. And if this regress is pursued, we must eventually come to beliefs for which there are not even Kornblithian reasons. This seems to show that while it is *possible* for an externalist to have Kornblithian reasons for his claims and even further Kornblithian reasons for the genuineness of his initial Kornblithian reasons, the whole structure is still, from a critical standpoint, indiscernible from one in which none of the reasons are genuine (*ibid* 332).

Notice the similarity between Bonjour’s point and Kaplan’s discussed above. Like Kaplan, on Bonjour’s view, a case in which Kornblithian reasons are present and one in which they are absent may look indiscernible to the agent. From the Impatient Cartesian’s perspective the ESI project can only be undergone by reflecting on reasons, and using them (if they exist) to justify a belief. Kornblithian reasons, argues Bonjour, are not the kinds of reasons that can be put to work in this way, since the agent will often not know whether she has them. Hence from Bonjour’s perspective, Kornblithian reasons are not useful for assessing or improving an agent’s epistemic situation. Again, this is because Bonjour thinks that Kornblithian reasons cannot be critically assessed either by one asking for a reason or by the person offering it himself. We can have Kornblithian reasons for many of our beliefs, and not have them for many others, and from the point of view of the agent, these beliefs may look the same. Any kind of *reasons* for belief that have such a strange feature, according to Bonjour, should not count as reasons at all. Simply

put, we can have these kinds of reasons, and still not be able to tell whether our beliefs are likely to be true.

Now suppose that the belief in question is the claim that reflection is unreliable for the purposes of error detection. Such a belief is inferential, and as such, seems to require rational defense in order to be justified. On Kornblith's view Kornblithian reasons will suffice to justify this belief, but nothing about Kornblithian reasons seem to give the agent any indication that the belief is true. Bonjour's objection might be summarized as follows: While Kornblith's critique of the Impatient Cartesian use of reflection in implementing the ESI project centers around the charge that it does not get the agent any closer to truth, his own view of justification as reliability preserves the connection between justification and truth at the cost of not being able to tell justified beliefs from ones that are not. Thus when asked about the reliability of the inference from the premises of the argument to its conclusion, Kornblith cannot give the kinds of reasons needed to accept the inference without retreating to internalism. This makes Kornblith's view look self-undermining.

§ 5. An externalist reply

Kornblith has acknowledged Bonjour's objection and seems unconvinced by it. Furthermore I think that Kornblith's response to Bonjour has significant import to Kaplan's objection. Finally, I think there is a response available that fits for both of the objections. This section will look at each of these in turn.

5.1 Kornblith's response to Bonjour:

In his response to Bonjour, Kornblith raises the following example. Suppose that S believes that P, and suppose that P is reliably produced, but that S does not have any internalist reasons for believing that P: S has no reflective access to the process by which P was produced and has no reflective access to whether the process which produced P is reliable. (Kornblith, 2006, 346-347). As should be clear from above, Bonjour will say that S must not have any genuine reasons for believing P, and this is clear given that without reasons that can be discovered by the appropriate kind of reflective access (the ones ruled out *ex hypothesi*) S can't tell whether P is true.

Kornblith's response is that he doesn't see why "being able to tell whether p is true" requires having this kind of reflective access to the reliability of the process by which the belief is produced. I can tell whether P is true if I am appropriately responsive to situations in which P is true. Nothing more than that is required" (*ibid* 347). A little further down he writes,

“Kornblithian reasons require reliable connection with the world. I can understand why I would want that, and what would be missing if I lacked it. There seems to be an important distinction between people who do, and those who do not, have this kind of connection to the world” (*ibid* 347). Given BonJour’s commitment to internalist reasons, he would suggest that in the imagined case where S believes that P in the absence of good internalist-brand reasons for P, S simply ‘cannot tell’ whether P is true. Kornblith’s suggestion here seems to be that to say that one simply ‘cannot tell’ whether P is not quite right. In fact, Kornblith seems to suggest that there might be a very straightforward way in which S *can* tell that P is true. S is usually responsive to situations in which the world is such that P is the case, and S has the appropriate connection to the world in order to determine that P. In such a situation, S forms the belief that P because S can tell that P is true. Nothing more than that seems to be needed for S to be justified in believing that P.

5.2 What Kornblith’s reply to BonJour means for Kaplan’s objection

Given Kornblith’s response to BonJour, a number of points carry over to Kaplan’s objection. First, Kaplan claims that naturalists are committed to the following conditional:

If “S is justified in believing P” means nothing more than “S formed P by a reliable process” [and it is not required of S that S be able to explain the causal origin of her belief that P for P to be justified], then S will be unable to tell which beliefs S ought to divest herself of, and which beliefs she ought to maintain.

Kaplan further suggests that few naturalists would be worried by being committed to this. I think, however, that Kornblith would not accept the truth of the conditional. Rather, Kornblith would say here, just as in his reply to BonJour, that there is a sense in which the agent *can* tell the difference between reliably produced and unreliably produced beliefs. There are many environments in which agents are responsive to the reliability of the processes at work in belief formation. Perhaps Kornblith would accept a revised version of the conditional, changing the consequent to read “S will be unable to tell *using reflection alone*, which beliefs S ought to divest herself of, and which beliefs she ought to maintain.” But this revised version seems to take the thunder out of Kaplan’s point.

Additionally, I think that Kornblith would further reject that he is committed to claiming that his theory is nonregulative in the way that Kaplan describes it. While I think he would maintain that it is true that an agent is not required to know the justificatory status of belief for that belief to be justified, this is not to say that he offers nothing in the way of belief regulation. As seen above (2.3), there is a perfectly cogent way in which Kornblith does offer epistemic advice.

5.3 Who's being self-undermining here?

The common thread between the objections from Kaplan and BonJour seems to be that if we give up the practice of reflecting on our reasons for a given belief as a means to justify it, we as agents will be unable to tell the difference between situations in which our beliefs are justified and when they are not. Further, by giving this up, as recipients of Kornblith's arguments, we will have no way to assess those arguments without getting some old-fashioned internalist brand reasons for accepting that argument. Hence to give up reflection and reason-giving as a means to justify a given belief—here the belief that reflection is unreliable for the purposes of error detection—is to give up our only means by which to judge the argument to be a good one. This, as I have said, makes Kornblith's position seem self-undermining.

One immediate problem with this kind of criticism might be seen if we consider how this debate has been framed in the first place. First, as I mentioned above, the intuition that Kaplan [and it seems BonJour] is demanding be preserved is INT1: If S is justified in believing that P then S, can produce upon reflection, her reasons for believing P. But a significant part of Kornblith's point is in showing that the kinds of reasons needed to preserve this intuition fail to materialize in many important cases. Again, Kornblith does not make this point by providing some *a priori* principle that opposes the intuition, but with empirical evidence to the contrary. While there may be a lot of common sense prima facie plausibility in INT1, the data suggest we drop it. The result of an agent continuing to cling to INT1 despite this contrary evidence, as Kornblith puts it, is “a false sense of confidence in an already misguided agent” (2002, p.115). To demand that such an intuition be preserved, and to demand that internalist-style reasons be used justify a belief, and to further demand Kornblith to give these kinds of reasons in favor of his argument is to demand Kornblith to do the very thing he has just shown not to work. One way to put this point is by saying, “Look, I just presented you with an argument that shows that the kinds of reasons you are demanding for the justification of the inference involved in making that argument fail to justify. If I were to attempt to give you these kinds of reasons, there would be good reason to think that they wouldn't justify the inference anyway.”

Kaplan and BonJour both claim that reliabilist justification is not reason providing, and therefore that it provides no grounds for evaluation, nothing by which to engage in the ESI project. But notice that the standard of evaluation that Kaplan and BonJour demand is an internalist standard. There might be perfectly good Kornblithian reasons for rejecting AOP, reasons which need not be grounded in that thesis. Likewise in arguing against the reliability of reflection for picking out problematic beliefs, there might be perfectly good Kornblithian reasons.

BonJour's worry is that while both Kornblith's claims may come out true, there are no internalist-style reasons available to justify them. This again, ought not to be a concern for Kornblith, or any other externalist who might make such claims. This charge only works if "justification" is taken to mean something other than what Kornblith means by it. It may be true in each case is that Kornblith is free from the responsibility of offering up internalist reasons for these claims, but why should this be a problem for him? While an internalist might think himself required to give such reasons for justification, Kornblith's whole point is that such reasons do not give a kind of justification that is desirable. Why fall into the same trap he has just shown the internalist to fall into? So, one might argue, it is not self-defeating to argue as Kornblith does. What would be self-defeating, however, would be to accept the demands of the internalist to provide their brand of reasons for this line of argumentation.

This still leaves the question of argument evaluability. While it seems that Kornblith is free from the worry of having an inconsistency in his theory and practice—offering what appear to be the kind of arguments in support of his view that his view dubs as unsatisfactory for justificatory purposes—then what kind of method of evaluability ought to be used? It does not seem to me that the only available method of evaluating an argument is one based on internalist-style reason giving. Rather, it seems that there is a perfectly coherent way to have a notion of good argument that does not rely on such faulty mechanisms. In claiming that reflection is unreliable at detecting errors in our beliefs Kornblith does not completely reject *all* notions of reflection as unreliable. To do so would be a collapse to skepticism. Rather, what seems to be the case here is that he is warning his audience that simply to rely on reflection as a justifier will not make our beliefs better connected to the world. This does not mean that one cannot give an argument to that end; it simply means that whether an inference to a given belief from other beliefs is a good one depends on the reliability of the process. This means getting out of one's armchair and examining some scientific data. As is the case for any other belief P, adding beliefs *about* the reliability of the process by which P was formed does not make P more or less reliable itself. At best it expands the number of beliefs one has in general.

§ 6. Reason Alienation:

If what has been put forth in the last section is correct, then Kornblith needn't worry about the charge of being self-undermining either in his critique of the system used in the Impatient Cartesian ESI project (AOP), or in his critique of its mode of implementation (reflection). This does not mean, however, that none of what has been said about the Kornblithian view of the ESI project is illuminating. Above I mentioned that while I thought that

the conclusion that Kaplan and Bonjour each draw about Kornblith's view is off base, something about their objections seems right. In this last section, I will attempt to figure out and explain what feature of Kornblith's view these objections touch on. I will leave it an open question the extent to which this is feature is a problem.

Given the complexity of what I am trying to get a handle on in this section, namely a phenomenon I call *reason alienation*, rather than beginning with an attempt to give some kind of definition or statement of it, I will begin with an anecdote describing the case in which I first noticed this feature of Kornblith's view. I think in describing this base case, a picture will emerge that can then be brought into tighter focus.

Similar to Bonjour, I noticed something that seemed like tension in Kornblith's methodology in chapter four of his (2002) book. Consider again, then, Kornblith's argument against the reliability of reflection at detecting errors in our beliefs (given in section 3.2), and his alternative epistemic advice, KEA. A pedantic little summary of the argument might be given as follows.

1. It is terribly difficult for an agent to use reflection to detect unreliably produced beliefs.
 2. When an agent reflects on an unreliably produced belief, the reasons that agent comes up with would be good if they were really the agent's reasons, but they are not.
 3. If 1 and 2, then reflection is an unreliable process for error detection.
- Therefore,
4. Reflection is an unreliable process for the purpose of error detection.

Suppose that from what I can tell, I take a look at this argument, and I decide to form the belief that 4; as a result I decide to use KEA as a means to improve my epistemic situation. I then realize that, along with having formed the belief that 4, I have also formed the belief that I formed the belief that 4 on the basis of (1-3). This second belief is my explanation for why I believe 4; or at least, it is *an* explanation for why I believe that 4. It is safe to call it my explanation if I am justified in believing that I formed the belief that 4 on the basis of the argument.²² But, notice what the contents of the premises are really saying. Premises 1-3 themselves tell me that, in general, 1-3 may be my reasons for believing 4, but they need not be. I might be just like Albert in Kornblith's critique of AOP. Now, how do I come to rest assured that 1-3 are my reasons for believing 4? Perhaps I just like the sound of 4; perhaps I have an internalist friend with whom I just love to argue, and toward whom I enjoy being contrary. I don't *think* that this is why I

²² Here I am taking justification to mean reliability.

believe 4, but for the very reasons supplied by the argument, I cannot tell using reflection whether the argument is really why I have decided to believe 4. Thus it seems that I am alienated from my reasons for believing the conclusion of the argument. The premises might be my reasons, and the argument itself is likely a reliable process, but I cannot be sure that this was the causal process responsible for *my* believing that 4. Nor does it seem that there will be any sort of scientific evidence available to help me analyze myself from a third-person perspective. There has never been an experiment conducted, so far as I know, that confirms the hypothesis “Jayme believes that 4 on the basis of 1-3.” Now I do know, given our current assumptions, that if my belief that I formed the belief that 4 on the basis of 1-3 is justified that it is likely to be true. But while I am sensitive to a great many environments in which I tend to form beliefs reliably, it is not clear that this is one to which I am sensitive. If it were, there would be a sense in which I can tell, just as Kornblith suggests, but given the complexity of such beliefs (beliefs about whether I have formed another belief on the basis of an argument, or on the basis of some other unreliable process).

My uncertainty does not stop there. Upon accepting 4, I also accept that many of the beliefs I think are justified may not be, and that reflection will not help me to pick these out. The whole problem, as I put it above, is that my bad beliefs look just as good to me as my good ones. The good ones that I can tell are good will most likely be produced in environments in which I am responsive to my own reliability in these environments. But there will be some good ones that were produced by reliable processes which I am not confident are reliable for me, and some bad ones that by definition will be unreliably produced. Well-formed beliefs about which I ‘can tell’ in the Kornblithian sense are not the ones I need to worry about. It is the good ones that I can’t tell are good that should trouble me, and the above argument tells me why: they look to me to be the same as the bad ones. Simply put, in carrying out KEA, one belief that should be at the forefront of my mind is that I don’t, and sometimes can’t, directly know what my reasons are for believing something, no matter how badly I may desire know this. This phenomena, again, is best described as reason alienation.

I want to get clear that this worry is not a matter of justification. By believing 4, and accepting KEA, I accept that regardless of what I may know, or think I know about a given belief, it is justified if it is reliably produced. Hence it should be made clear that I can be justified in believing P, and still be alienated from my reasons for P. In fact, one might say that the only way there will be reasons to be alienated from at all is if such reasons, in fact, exist. If the belief is unreliably produced, there will be no reasons to be alienated from. So perhaps a better way to put it is that if I am alienated from my reasons for P, then such reasons, in fact, exist (it just isn’t clear to me what they are). Notice, then, that unlike Bonjour and Kaplan, I am not demanding that the

externalist apply an internalist standard of justification to her theoretical beliefs. *Those* kinds of reasons, by themselves, would be no help, as Kornblith has aptly shown.

While this is all still very sketchy, perhaps enough has been said to get a flavor for the kind of environment in which reason alienation occurs. First, one needs to believe that reflection is unreliable for the purposes of error detection. Then, one embarks on the Kornblithian understanding of the ESI project: one accepts KEA as good epistemic advice. Under such conditions, while some of the beliefs I have will be reliably produced, and I will be able to tell in the Kornblithian sense of ‘able to tell’ that they are true, for others I will not be able to tell so easily, or at all. Beliefs of these latter types may be quite complex—e.g. beliefs about what my actual reasons for some belief are—and very difficult to evaluate even with the help of cognitive psychology. While it may not affect the justificatory status of these beliefs that I may never know for certain what my actual reasons are for having them, it does affect my psychology. The realization that, reflectively, these beliefs look to me the same as the bad ones, will make me distrust some of my beliefs that I am perfectly justified in believing. Hence, it seems that having the experience of being alienated from one’s reasons is integrally tied to following the externalist epistemic advice.

Above I mentioned that I would leave it an open question the extent to which reason alienation was a problem for the externalist understanding of the ESI project. Part of the reason for my leaving this question open is that I suspect that whether it is viewed as a problem or not will have a lot to do with one’s previously held intuitions about the plausibility of externalism. Perhaps an analogous ‘problem’ for the internalist conceptions of justification is that internalism is committed to divorcing justification from truth. When it comes to understanding what the ESI project consists in it seems we are met with a choice of what seems most important to us. The choice is between whether we want to follow the internalist, confidently call many of our beliefs justified, but have that claim mean little in terms of the truth of that belief, or follow externalist epistemic advice, know that a justified belief is very likely to be true, but accept that there will be many beliefs whose reasons from which we are alienated.

References

BonJour, Lawrence. "Kornblith on Knowledge and Epistemology" *Philosophical Studies* (2006) 127:317-335.

Elga, Adam. "On Overrating Oneself...and Knowing it" revised MS, December 10, 2003, to appear in *Philosophical Studies*.

Feldman, Richard. "Chisholm's Internalism and Its Consequences," *Metaphilosophy* Vol. 34, No.5, (October 2003) 603-620.

Kaplan, Mark. "Epistemology Denatured," *Midwest Studies in Philosophy*, XIX (1994) 350-365.

Kornblith, Hilary. "Reply to Bermudez and BonJour" *Philosophical Studies* (2006) 127: 337-349.

— "Can Internalism Be Saved?" *Metaphilosophy* Vol. 34, No.5, (October 2003) 621-629.

— "Roderick Chisholm and the Shaping of American Epistemology," *Metaphilosophy* Vol. 34, No. 5, (October 2003) 582-602.

— *Knowledge and its Place in Nature*. Oxford University Press, Oxford (2002).

— "Distrusting Reason," *Midwest Studies in Philosophy*. XXIII (1999) 181-196.