



Loeb on Stability and Justification in Hume's *Treatise*

Frederick F. Schmitt

Hume Studies Volume 30, Number 2, November (2004) 297-328.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact

humestudies-info@humesociety.org

<http://www.humesociety.org/hs/>

A Symposium on Louis E. Loeb,
Stability and Justification in Hume's Treatise

Loeb on Stability and Justification in Hume's *Treatise*

FREDERICK F. SCHMITT

In *Stability and Justification in Hume's Treatise*,¹ Louis Loeb ascribes to Hume a naturalistic account of justified belief, one on which Hume is fundamentally concerned with the question whether stable belief can be achieved. Loeb's interpretation is systematic, richly explanatory, and powerfully argued. He makes a compelling case that stability plays a central role in Hume's epistemology. Loeb's case is so compelling indeed that anyone who wants to defend an alternative interpretation will now have to assimilate or deflect the massive textual evidence in favor of the stability interpretation. I will argue here that, for some passages Loeb cites in favor of the stability interpretation, a veritistic interpretation explains the text at least as well as the stability interpretation does.

1. The Stability Account of Belief and the Aim of Truth

On Loeb's interpretation, a belief is by its nature a steady or infixed disposition to thought, will, passion, and action—i.e., a disposition steady in its influence on thought, will, etc. (SJ 65–74). The nature of belief derives from its natural function, described by Hume at T 1.3.10.2 (SBN 119):²

Did impressions alone influence the will, we shou'd every moment of our lives be subject to the greatest calamities; because, tho' we foresaw their approach, we shou'd not be provided by nature with any principle of action, which might impel us to avoid them. On the other hand, did every

idea influence our actions, our condition wou'd not be much mended. For such is the unsteadiness and activity of thought, that the images of every thing, especially of goods and evils, are always wandering in the mind; and were it mov'd by every idle conception of this kind, it wou'd never enjoy a moment's peace and tranquillity.

If impressions but not ideas influenced the will, our ideas of the future would provide no guidance in avoiding calamities or evils (or, I take it from the third sentence quoted, in achieving goods). But if every idea influenced the will, then, since ideas of all kinds of objects wander in the mind, the mind would enjoy no rest. Nature accordingly deprives "images" of a significant or regular influence on the will and reserves that influence for impressions and ideas of a special kind, so that we tend to act only when guided either by impressions or by such ideas. These ideas of a special kind must not waver and thereby deprive us of rest. The passage suggests that there is a natural kind of idea that serves the desired natural function. I take it that, on Loeb's interpretation, the kind belief is simply identified with this special kind of idea (or more accurately, disposition). That is, beliefs are, by nature, ideas that serve this natural function. To serve this function, beliefs must have a steady influence on thought and the will. A belief, then, is a disposition steady in its influence on thought and the will (with a qualification for countervailing influences to be mentioned below). Loeb argues persuasively that this stability interpretation of belief makes sense of a range of textual phenomena in Book 1.

While Loeb convinces me that for Hume steady influence is *part* of the natural function and nature of belief, I find less plausible his assumption that it is the whole story. In interpreting Hume's account of belief, Loeb focuses exclusively on the point that beliefs differ from images in their function of a steady influence on will, allowing the mind a "peace and tranquillity" that would be disrupted by the influence of inconstant ideas on the will.³ If every image had a motivating effect, our will would turn every which way, with an attendant mental agitation. This is a purely psychological function of belief and an important one for Hume. But a close look at the passage reveals that Hume has much more in mind than the function of psychological tranquility. A system in which only impressions influence the will is deficient because impressions do not provide guidance that enables us to avoid calamities or obtain goods. For such guidance, we need ideas of the future. Such ideas must influence the will. The natural function of these ideas is to enable us to obtain goods and avoid evils. This is an *extra*-psychological function of the ideas. This extra-psychological natural function of belief is just as prominent in the passage as the function of affording tranquility. Thus, there are two components in the natural function of belief—steady influence on the will and guidance of the will by foresight. Loeb omits the second natural function. Loeb could maintain his interpretation of the nature of belief as a disposition with steady influence on the

will if he were able to impute one natural function, that of steady influence, to the nature of belief while denying that the other natural function, guidance, belongs to the nature of belief. But I see no basis, in this passage at least, for discriminating between the two natural functions in this way.

In describing what beliefs must be like to fulfill their guidance function, Hume uses the term “foresaw.” Taken literally, this would ascribe an *epistemically* defined natural function to belief, and indeed one that entails truth. Assuming again that the natural functions of belief belong to its nature, it would follow that beliefs are by nature true. Yet Hume certainly does not think that beliefs are always true. So he cannot mean “foresaw” literally, and perhaps he does not mean it epistemically either. The best interpretation is that he uses the word to suggest the condition he really does need: that the action-guiding ideas of the future are true often enough to allow us to obtain goods and avoid evils. On this interpretation, the natural function of belief, and hence its nature, entails that most of the guiding beliefs are true representations of the future. This entailment of a veritistic principle of charity regarding the content of belief is at odds with a pure stability account of the natural function and nature of belief.

Not only is guidance a natural function of belief in the quoted passage, but it may be a more basic natural function than steady influence. The function of steady influence may partly derive from that of guidance. Although Hume cites only the mind’s “peace and tranquillity” as an objective that rules out an influence by “every idea,” it is obvious from the first sentence of the quoted passage that such an influence would cause far greater harm than mere mental turbulence: it would lead perpetually to the pursuit of chimaeras (i.e., the pursuit of distant or imaginary goods, and the attempt to avoid such evils) at the expense of obtaining real and present goods and avoiding such evils. That Hume intends to refer to the latter and greater cost of the influence of every idea in the third sentence of the passage, and not merely to mental turbulence, is indicated by his qualification “especially of goods and evils.” I do not think that this qualification could be intended merely to suggest that ideas of goods and evils cause more mental turbulence than other ideas do. That would reconcile the qualification with a reading of the third sentence on which Hume intends to refer here only to mental turbulence and to no greater harm. But Hume cannot simply assume here that ideas of goods and evils cause more turbulence than other ideas do. For just how much ideas of any particular kind agitate us is up for stipulation in Hume’s thought experiment in the passage. The better reading of the qualification “especially of goods and evils” is that it refers back to the calamities mentioned in the first sentence. The pursuit of distant or imaginary goods wastes effort that might be used to obtain real and present goods, and it exposes us to unnecessary risks. This leads me to think that for Hume the natural function of steady influence at least partly derives from that of guidance. What is fundamentally valuable is for ideas to have an influence lead-

ing to actions that obtain goods and avoid evils. A steady influence is instrumental to this fundamental value. Most goods and evils endure for a period of time, and obtaining or avoiding them requires elaborate preparation. This makes it desirable and even necessary for success in action that the mind keep these goods and evils persistently or periodically in view when planning or taking action. This amounts to a steady influence on action. In addition to this point, successful *habits* of action require ideas of goods and evils to have a regular influence on the will. I am inclined, then, to think that steady influence is only part of the natural function of belief, at least partly derivative from guidance. Hume is not a pure stability theorist of belief.

Let us turn now to a philosophical objection to the pure stability account of belief. It is a commonplace that belief aims at truth. It is plausible that this is so by the very nature of belief. On the face of it, a stability account is inconsistent with this claim; for it makes no mention of the aim of truth. Loeb responds, to the contrary, that “considerations of stability absorb considerations of truth” (SJ 84). To develop this response, Loeb distinguishes two sorts of dispositions associated with a belief. There is a first-order disposition to influence thought, the will, and action. And there is “a second-order disposition to regulate one’s belief that a proposition is true, and hence the first-order dispositions involved in belief, by (what one takes to be) evidence or indicators of truth” (SJ 82). Having a second-order regulative disposition of this sort amounts to “the aim that belief should be sensitive to what is really true” (SJ 82). The claim that belief aims at truth is thus captured by the idea that in having a belief, the subject has a second-order regulative disposition of this sort. Loeb argues that for Hume subjects have a certain regulative disposition along these lines, citing Hume’s remark at T 1.4.1.6 (SBN 183): “When I reflect on the natural fallibility of my judgment, I have less confidence in my opinions.” Here Hume is committed to the assumption that the disposition to reduce confidence in a belief given reflection on the fallibility of one’s judgment is present in every case of belief—an assumption consistent with the view that it is part of having a belief to have a disposition to reduce one’s confidence in the belief, given such a reflection. The passage is best explained by attributing to Hume the view that in having a belief, I have a second-order regulative disposition to reduce confidence in my belief, given reflection on the fallibility of my beliefs.

All this seems right. Suppose Hume does accept that in having a belief, I have such a second-order regulative disposition. Even so, it does not follow that for Hume considerations of stability absorb considerations of truth. The question remains whether for Hume stability provides a *rationale* for the requirement of a second-order regulative disposition of the relevant sort in believing. A rationale of this sort might entail either of two conditions. It might entail the stronger condition that aiming at truth, as captured by the regulative disposition, contributes to *actual* stability. Or it might entail the weaker condition that aiming at truth (i.e., the

regulative disposition) contributes to *aiming* at stability, or to *expected* stability, in the sense that the disposition can be *reasonably expected* to contribute to stability. I am not sure which of these conditions must be met if considerations of stability are to absorb considerations of truth. Accordingly let me consider them in turn.

Regarding the stronger condition that the regulative disposition contributes to actual stability, it is clear that for Hume this condition is not satisfied. For the regulative disposition at T 1.4.1.6 (SBN 183) does not contribute to actual stability. The first point to make is that the disposition is merely conditional. It is a disposition to reduce my confidence in my belief upon reflection on the fallibility of my judgment. This disposition is not triggered in unreflective subjects. So it has *no* effect in those subjects. In the case of reflective subjects, in which it is triggered, the disposition has a *destabilizing* effect. (Loeb of course recognizes all this and gives an interesting account [85–6] of how the destabilization occurs.) The disposition does not lead to convergence on a fixed degree of confidence, but causes vacillation. So the disposition does not contribute to actual stability in either reflective or unreflective subjects. Granted, if the disposition did lead to the elimination of belief in a reflective subject, the resulting system of belief might well be more stable than the initial system. But the disposition in fact leads to vacillation rather than elimination.

Regarding the weaker condition, that the regulative disposition can be reasonably expected to contribute to stability, it is less clear what to say. A person as reflective as Hume will, in Hume's view, reasonably come to the conclusion that the disposition cannot be expected to contribute to stability, by the reasoning that leads to the conclusion that the disposition is in fact destabilizing in reflective subjects and neutral in unreflective subjects. A person less reflective than Hume might come reasonably but mistakenly to the conclusion that the disposition can be expected to contribute to stability. Then again, such a person might come to Hume's own view, or might come to no view at all. Thus, we are left not knowing whether the weaker condition is often satisfied for unreflective subjects. I conclude, then, that we have no basis for thinking that for Hume either the stronger or the weaker condition is often satisfied. There is no reason to think that for Hume stability provides a rationale for the requirement of the second-order regulative disposition in believing. So there is no reason to think that for Hume considerations of stability absorb considerations of truth.

Let me add two caveats to these points. I do not deny that for Hume certain kinds of reflection do contribute to actual stability. I will suggest in section 3 that Hume may attribute such a role to the Humean philosophical reflection involved in recognizing the "manifest contradiction" and the "dangerous dilemma" of 1.4.7. Such reflection tends to produce everyday beliefs in a reflective person that are more stable than those that would be produced without it. But there is no reason to think that for Hume a disposition to stable belief under Humean philosophical

reflection is present in *all* cases of belief. Humean philosophical reflection is more sophisticated than mere reflection on the fallibility of my judgment. There is no reason to think that such sophisticated reflection would contribute to the stability of the belief if it were triggered in a subject unaccustomed to sophisticated reflection. Humean philosophical reflection might well have a destabilizing effect on an unsophisticated subject. So the stabilizing effect of Humean philosophical reflection does not plausibly underwrite a regulative disposition, present in all instances of belief, that when triggered routinely contributes positively to actual stability.

My second caveat is that I do not deny that there may be some second-order disposition to regulate belief with respect to truth, present in all instances of belief that when triggered routinely contributes positively to actual stability. It may be that in every instance of belief the subject has a conditional disposition to adjust the belief upon reflecting on his or her intellectual virtues, and this disposition, when triggered, tends to stabilize the belief. My point is only that the regulative disposition at T 1.4.1.6 (SBN 183) does not do this, and I see no other evidence that for Hume any second-order disposition to regulate belief with respect to truth that is present in all cases of belief does contribute positively to actual stability. This leaves us without a clear case that stability provides a rationale for the requirement of a veritistic regulative disposition in believing, and thus without a case that for Hume considerations of stability absorb considerations of truth.⁴

2. The Textual Basis for a Stability Interpretation of Justified Belief

According to Loeb, Hume holds a stability account, not only of belief, but of *justified* belief: a belief is justified if it results from a mechanism that tends to produce ideas (or more exactly, dispositions) steady in their influence on the will. Hume develops this account in two stages (SJ 12–20). In the *positive* stage, Hume displays his pre-theoretical intuitions about justified belief and embraces the stability account as fitting these intuitions better than alternative accounts (1.3, 1.4.1–6). These intuitions include affirmative evaluations of beliefs based on causal inference and negative evaluations of the vulgar and philosophical beliefs in body (1.4.2) and the beliefs in material (1.4.3) and immaterial (1.4.5) substances. In the *negative* stage of his epistemology (1.4.1, 1.4.7), Hume concludes that the stability account does not vindicate his positive pre-theoretical intuitions about causal inference after all, at least not for a reflective subject like himself.

Loeb supports the stability interpretation of justified belief by appeal to the fact that Hume often makes psychological and epistemic claims in tandem; Hume passes without notice from talk of belief-formation to talk of justified belief (SJ 60–5; cf. 102). For Hume, “establishing that the states produced by a psychological mechanism are beliefs is to establish that they are justified” (SJ 65). Loeb’s explanation for this connection between belief-formation and justified belief is that for

Hume a justified belief is merely a belief that results from a belief-forming mechanism (i.e., a mechanism that tends to produce ideas that are beliefs—are steady in their influence). Any belief is infixed, or steady in its influence, in so far as it is produced by a belief-forming mechanism. But a “belief might fail to be steady in its influence owing to the presence of beliefs with which it conflicts, beliefs that reduce the likelihood of the occurrence of its characteristic manifestations or its typical effects, which reduce its influence on the will and action” (SJ 80). This leads Loeb to attribute to Hume a distinction between a belief justified *other things equal* and a belief justified *all things considered*. A belief is justified other things equal if it is infixed by a belief-forming mechanism. This is consistent with the belief’s failing to be steady in its influence owing to the presence of conflicting beliefs. A belief is justified all things considered if it is not merely infixed by a belief-forming mechanism, but steady in its influence even taking into account the effect of all other cognitive mechanisms that affect its steadiness.

This, in outline, is the stability interpretation of justified belief. We should now turn to its textual basis. Do Hume’s repeated “tandem” references to belief and justified belief support the stability interpretation of justified belief? I concede that they cumulatively provide prima facie evidence for the stability interpretation. Nevertheless, I want to urge that we can do without the stability interpretation in one location extensively discussed by Loeb—the sixth paragraph of T 1.3.9 (SBN 109). In this passage, Hume makes a case that fictions resulting from associations by resemblance and contiguity are not beliefs. The passage is this:

For where upon the appearance of an impression we not only feign another object, but likewise arbitrarily, and of our mere good-will and pleasure give it a particular relation to the impression, this can have but a small effect upon the mind; nor is there any reason, why, upon the return of the same impression, we shou’d be determin’d to place the same object in the same relation to it. There is no manner of necessity for the mind to feign any resembling contiguous objects; and if it feigns such, there is as little necessity for it always to confine itself to the same, without any difference or variation. And indeed such a fiction is founded on so little reason, that nothing but pure *caprice* can determine the mind to form it; and that principle being fluctuating and uncertain, ‘tis impossible it can ever operate with any considerable degree of force and constancy. The mind foresees and anticipates the change; and even from the very first instant feels the looseness of its actions, and the weak hold it has of its objects.

In this passage there is a tandem reference connecting a fiction “founded on so little reason” with one determined by “caprice.” Loeb plausibly reads “founded

on so little reason” as an epistemic expression meaning something like “so very unjustified,” rather than as a psychological predicate denying formation by reason.⁵ So Hume infers from the fact that such a fiction is so very unjustified that it is formed by caprice, an unstable mechanism (i.e., one that tends to produce unstable ideas). Justification is linked here with instability. One might think that the stability account plays a role in Hume’s inference here, since the inference moves from the unjustifiedness of the fiction to the instability of the fiction.

On close inspection, however, the passage leaves no room for Hume to employ the stability account. The ultimate aim of the passage is to argue that fictions resulting from associations by resemblance and contiguity are not beliefs. This conclusion is supported on the ground that such fictions are unstable: “There is no manner of necessity for the mind to feign any resembling contiguous objects.” The sentence in which the tandem reference occurs is clearly intended as supplementary support for this ultimate conclusion, as indicated by Hume’s prefatory use of the word “indeed.” The tandem sentence is intended to show that fictions resulting from associations by resemblance and contiguity are unstable. Their instability is inferred from the claim that they are determined by caprice, on the ground that caprice is an unstable mechanism. On the face of it, the claim that caprice is unstable is introduced as an unsupported premise (“and that principle being fluctuating and uncertain”), not inferred from the fact that caprice forms the unjustified fiction. Hume infers that the fiction is determined by caprice on the ground that it is “founded on so little reason.” Putting all this together, Hume’s argument in the tandem sentence is:

A fiction resulting from association by resemblance or contiguity is so very unjustified.

Such an unjustified belief must be formed by a mechanism that is very unjustifying (where a justifying mechanism is one that produces beliefs justified other things equal).

Caprice is such a mechanism.

Hence, this fiction results from caprice.

Additional premise: Caprice is an unstable mechanism.

Hence, probably, this fiction is unstable.

If this is the right reading of the tandem sentence, then the stability account clearly plays no role in Hume’s inference.

I have said that, on the face of it, Hume does not infer that caprice is an unstable mechanism from the claim that it is unjustifying. If, however, Hume did make such an inference, the inference would certainly rely on the stability account, and the passage would support the stability interpretation. So it is important to determine whether Hume does or doesn’t make this inference. In addition to the

surface textual evidence against saying that he does, there are two further points. First, if Hume accepted the stability account, he could infer directly from the fact that the fiction is unjustified that it is likely to be unstable, since, on the stability account, most unjustified beliefs are unstable. There would be no need to bring caprice into the inference. But Hume does not make this inference. The fact that he does not do so is a tad of evidence *against* the stability interpretation. Second, the instability of caprice is an obvious phenomena—at least as obvious as its lack of justifying power. For this reason it would be otiose to appeal to the unjustifiedness of caprice to establish its instability. I conclude that the stability account plays no role in this tandem passage.

I would add to this that the passage at T 1.3.9.6 (SBN 109) contains minor evidence for a reliability interpretation.⁶ Hume observes that associations of resemblance and contiguity give rise to ideas ascribing different and varying relations to objects from one occasion to the next. This observation occurs in the sentence immediately preceding the tandem sentence. The order of discussion suggests that the observation is intended to *support* the premise at the beginning of the tandem sentence (that “such a fiction is founded on so little reason”). Certainly Hume’s argument in the tandem sentence would be stronger if his premise received such support. One might think that a stability interpretation could explain how the observation supports the premise, as follows. Hume observes that associations of resemblance and contiguity give rise to ideas ascribing different and varying relations from one occasion to the next, hence yield unstable ideas. From this conclusion of instability, Hume infers, via the stability account of justified belief, that the fictions these associations yield are unjustified—“founded on so little reason.” But there is an obstacle to this stability-based explanation of how Hume’s observation supports the premise that fictions are “founded on so little reason”: in the remainder of the tandem sentence Hume infers the capriciousness of the ideas produced by the associations, and hence their instability, from their unjustifiedness. Thus, inferring the unjustifiedness of the ideas produced by the associations from their instability would introduce a tight circularity into Hume’s argument in the tandem sentence. I conclude that the stability account has no role to play in supporting the premise that the fiction is founded on so little reason.

The better reading of T 1.3.9.6 (SBN 109) is that Hume offers two independent arguments for the conclusion that these fictions are unstable. Both arguments appeal to the observation that the associations of resemblance and contiguity give rise to ideas ascribing different and varying relations. In the first argument, Hume infers from this observation that such ideas are unstable; so the fictions to which the associations give rise are unstable. In the second argument, Hume infers from the same observation that these fictions are so very unjustified, hence result from a very unjustifying mechanism, caprice, and, being formed by a mechanism (caprice) that produces unstable ideas, are unstable.

A reliability interpretation can allow for these two independent arguments by attributing to Hume a deliberate ambiguity in the two sentences (T 1.3.9.6; SBN 109) claiming that the associations give rise to ideas ascribing different and varying relations from one occasion to the next. In these sentences, Hume proposes not only that the ideas from associations are unstable because they ascribe different and varying relations, but also (tacitly) that many are false because they ascribe different and varying relations, hence ascribe contrary relations and relations that vary when there is no variation in the ideas. Associations of resemblance and contiguity yield contrary ideas and are thus unreliable. From this tacit proposal that the ideas are often false, Hume infers that the fictions the associations yield are unjustified. Caprice is the only mechanism that produces ideas so unjustified. So these fictions are produced by caprice, hence unstable. The reliability interpretation has this to be said for it: it explains how Hume's appeal to caprice can genuinely and noncircularly supplement his argument that the associations give rise to ideas ascribing different and contrary relations and so are unstable. Thus, the reliability account of justified belief may play a role in the passage. A role for the stability account seems unlikely.

3. Actual Versus Reflective Stability Accounts of Justified Belief

Loeb distinguishes two stability accounts of justified belief (all things considered) that might be attributed to Hume (SJ 88). On what I will call the *reflective* stability account, a belief is justified for any subject (whether a reflective or unreflective person) if it results from a mechanism that would tend to produce stable beliefs were it used by a reflective person. On what I will call the *actual* stability account, a belief is justified for a subject if it results from a mechanism that tends to produce stable beliefs given the actual degree to which that subject is reflective. The reflective and actual stability accounts agree in their assessments of the justified beliefs of reflective persons. But they diverge in their assessments of the justified beliefs of unreflective persons. For causal inference is destabilized in the reflective but not the unreflective person (1.4.7), and thus the reflective stability account judges the beliefs of unreflective persons to be unjustified, while the actual stability account judges these beliefs to be justified.

Loeb prefers the actual stability interpretation to the reflective stability interpretation, for two reasons. First, there is a philosophical objection to the reflective stability interpretation: reflection is not instrumental to the stability of ideas; hence the value of stability provides no rationale for the reflective stability account (SJ 90n48, 92). The conclusion of Book 1 can be seen as casting doubt on whether reflection is instrumentally valuable for stability. Indeed, Hume argues to the contrary that reflection is destabilizing and so undermines the natural function of beliefs.

This objection to the reflective stability account seems to me basically right. Two remarks:

(i) There is some reason to think that for Hume, reflection is instrumental to stability. More exactly, it is instrumental to stability in those subjects who follow Hume through the “manifest contradiction” between causal inference and the continued existence of matter (T 1.4.7.4; SBN 266) and the “dangerous dilemma” between a false reason and none at all (T 1.4.7.6; SBN 267). To be sure, as I urged in section 1, ordinary reflection on the fallibility of one’s judgment destabilizes belief (1.4.1). And the Humean philosophical reflection that exposes the manifest contradiction and dangerous dilemma destabilizes reason in a reflective person *while that person reflects*. But it may be that such philosophical reflection in the end contributes to stable belief in everyday life, once the subject has ceased reflecting. A remark of Hume’s suggests that the reflective person who has gone through Hume’s own philosophical reasoning exposing the manifest contradiction and dangerous dilemma and has then moved from the study to everyday life will obey natural inclination in a manner different from that of the unreflective person: “in this blind submission [to my senses and understanding] I shew most perfectly my sceptical disposition and principles” (T 1.4.7.10; SBN 269). Among the views suggested by this remark is that the Humean reflective person will restrict beliefs to those formed by causal inference and avoid using other operations of imagination. By restricting his or her beliefs in this way, the Humean reflective person will form stable beliefs more often than an unreflective person will. An unreflective person will not stick to the senses and understanding but occasionally form unstable beliefs by fancy. If this is Hume’s suggestion, then reflection has instrumental value for the stability of the beliefs of Humean reflective people in everyday life.

(ii) Yet, even if all this is so, Loeb is still right to deny that the value of stability provides an adequate rationale for the reflective stability account of justified belief. One point here is that the reflective stability account ignores the stabilizing effects of reflection during everyday life and considers only the stability of beliefs *during* reflection. Thus, the fact that Humean reflection has a stabilizing effect on beliefs in everyday life provides no rationale for the reflective stability account. Another point is that the reflection to which the account refers generally falls short of sophisticated Humean philosophical reflection. The stabilizing effect of Humean reflection does not entail that less sophisticated reflection has a stabilizing effect. A final point, the most significant, is that the stabilizing effect of Humean (or for that matter, any other) reflection provides no rationale for imposing the requirements of reflection on *unreflective* subjects. Even if Humean philosophical reflection enhances stability in everyday life, it does so only for philosophers who engage in such reflection. The value of stability rather motivates an account of justification on which a subject is justified in a belief only when the belief results

from a mechanism that produces stable beliefs *in the subject*. Thus it motivates a condition that allows unreflective people to use mechanisms that produce stable ideas in everyday life. There is no rationale in stability for insisting that unreflective people are justified in a belief only if it results from a mechanism that would be stabilizing in reflective people during reflection (or afterward, for that matter). So, although reflection may have instrumental value for stability in reflective people in everyday life, Loeb is right that the value of stability offers no rationale for the reflective stability account.

Loeb's second, and I believe main, reason for preferring the actual to the reflective stability interpretation is textual (SJ 91–3). On the reflective stability interpretation, Hume's positive claim in part 3 that causal inference is justified is logically inconsistent with his claim in 1.4.7 that causal inference yields unstable beliefs in reflection. For according to the reflective stability account, the positive claim in part 3 is simply that causal inference would yield stable beliefs if used by a reflective person. And this is logically inconsistent with the negative claim of 1.4.7 that causal inference yields unstable beliefs in reflection. Thus the reflective stability interpretation forces us to see Hume in 1.4.7 as *retracting* his part 3 positive evaluation of causal inference in order to avoid logical inconsistency. This is *prima facie* a serious drawback of the interpretation, since there is no clear textual evidence that Hume intends in the Conclusion to retract his positive part 3 evaluation. Though the evaluations of part 3 and 1.4.7 are, on the face of it, inconsistent, Hume never lets on that he is retracting his position. The actual stability interpretation, by contrast, reconciles Hume's positive and negative evaluations of causal inference. On this interpretation, Hume is claiming in part 3 that for an unreflective subject causal inference is stable and justifying, while in 1.4.7 he is claiming, consistently with part 3, that causal inference is unstable and unjustifying for a reflective subject like himself. There is no inconsistency. In short, according to Loeb, the actual stability interpretation has the advantage over the reflective interpretation that it can reconcile Hume's positive evaluation of causal inference in part 3 with his negative evaluation in 1.4.7.

The first point to make in response to this suggestion is that it is not clear on the textual evidence that Hume's positive evaluation of causal inference in part 3 is intended to apply only to unreflective subjects, or that his negative evaluation in the Conclusion is intended to apply only to reflective subjects. Loeb addresses this very broad and difficult interpretive issue and offers a *prima facie* textual case in favor of his view. I will have to waive this issue here and rest content to observe that if, as I suspect, the positive and negative evaluations are supposed to apply to all subjects, then the actual stability interpretation has no advantage over the reflective stability interpretation. Both stability interpretations will run afoul of Hume's intentions: the actual stability interpretation will be at odds with the negative evaluation of unreflective subjects in 1.4.7, while the reflective stability

interpretation will be at odds with the positive evaluation of unreflective subjects in part 3. Of course, if Hume evaluates all subjects both positively and negatively, his claims will have to be reconciled by some nonlogical device—on which I will say more momentarily.

I propose to set aside this issue and grant for the sake of argument that Hume intends his positive part 3 evaluation to apply only to unreflective subjects and his negative 1.4.7 evaluation to apply only to reflective subjects. On this assumption, Loeb is certainly right that the actual stability interpretation can, and the reflective interpretation cannot, *logically* reconcile Hume's positive (part 3) and negative (1.4.7) evaluations of causal inference. This is an impressive advantage of the actual stability interpretation, given the assumption on which we are proceeding. But, impressive as it is, the advantage does not strike me as decisively favoring the actual stability interpretation. For it is possible to read Hume as retracting his positive part 3 evaluation when he comes to 1.4.7. Loeb recognizes that a proponent of the reflective interpretation can see Hume's positive part 3 evaluation as a provisional one, to be retracted in light of Hume's developing associationist cognitive psychology. On what I take to be the most plausible version of this *provisionalist* interpretation, Hume's evaluations of causal inference in part 3 and 1.4.7, though contradictory, are appropriate under proper philosophical methodology, which calls for making evaluations in the course of an inquiry in light of the psychology and philosophical judgments available at that moment in the inquiry. A proponent of the reflective stability interpretation may say that in light of the psychology and philosophy available in part 3, we would judge causal inference to be stable in a reflective person, and thus we would evaluate causal inference positively. But in light of the judgments available in 1.4.7, we would evaluate causal inference negatively. (I note that the reliability interpretation, like the reflective stability interpretation, must reconcile Hume's earlier and later evaluations—whether taken as restricted to unreflective or reflective subjects, respectively, or not—by something like the provisionalist reconciliation. For, on the reliability interpretation, the earlier and later evaluations are logically inconsistent with one another.)

I add here a doubt about whether either stability account, the actual or the reflective, has a rationale in the value of the stability of ideas. On both accounts, whether a mechanism is justifying for a *reflective* person turns on whether it yields stable ideas during reflection, not during everyday life. (To see this, recall that on the actual stability account, a belief is justified for a subject only if it results from a mechanism that produces stable beliefs under the *highest* level of reflection the subject actually achieves, and the highest level of reflection achieved by a *reflective* subject of course involves reflection.) But in my view, if a stability account is to be plausible, then whether an unreflective belief of a reflective person is justified should turn on whether the mechanism that produces it yields stable beliefs dur-

ing the unreflective moments of that person's life. This is true if we require the mechanism to yield stable beliefs on average over the subject's entire belief system.⁷ Such a requirement is motivated by the value of stability. What is not so motivated is requiring the mechanism to yield stable beliefs over the portion of the belief system (reflective) to which the given unreflective belief does not belong.⁸ This line of reasoning leads to the conclusion that, on a stability account, we should evaluate a mechanism for a reflective as well as an unreflective subject by looking at its stability over its total output in the subject. Plausibly, causal inference yields stable beliefs in its total output in both a reflective and an unreflective subject. And, though the identity-ascribing propensity yields the philosophical belief in body in a reflective subject in the study, it yields the vulgar belief in body in reflective subjects "upon leaving their closets" (T 1.4.2.53; SBN 216), so that it too yields stable beliefs over its total output in a reflective subject. If we value stability, we ought to require, for an unreflective belief of a reflective subject, a mechanism that produces a high average stability over its total output in the subject, not one that produces a high average stability over the subject's most reflective output. For this reason, neither the actual nor the reflective stability account is motivated by the value of stability.

Of course these points do not undermine a stability motivation for a variant of the actual stability account: that a belief is justified for a subject when it results from a mechanism that produces beliefs with high average stability over its total output in the subject. But, as I have observed, this stability account evaluates causal inference positively and is thus inconsistent with Hume's negative evaluation of causal inference in 1.4.7. Indeed, this observation generalizes to any stability account motivated by the value of stability: any such account will be inconsistent with Hume's negative evaluation of causal inference in reflection.

The reliability account is not motivated by the value of stability, but it can be motivated by the value of guidance in action, enabling us to obtain goods and avoid evils. It can at the same time respect Hume's negative reflective evaluation of causal inference. According to the reliability account, to evaluate whether causal inference is justifying is not to assess the effect of reflection on the outputs of causal inference. Thus, it is not to assess whether causal inference is justifying for a reflective subject. It is to evaluate theoretically, during reflection, a feature of causal inference that remains constant between everyday life and reflection—its reliability. Reflection bears on causal inference only in enabling us to evaluate causal inference systematically. What we judge, when in possession of full information, is that its reliability is at best uncertain. This is a worry about our capacity for obtaining goods and avoiding evils in everyday life. Thus, the reliability account differs from the stability account in having some chance of being motivated by Hume's account of the natural function of belief (on my interpretation of that function), while explaining Hume's negative evaluation of causal inference in 1.4.7.

4. The Stability Account and Philosophical Beliefs

Loeb supports the stability interpretation of justified belief by appeal to Hume's evaluation of various philosophically important beliefs—namely, the vulgar belief in body based on constancy (as opposed to coherence), the philosophical belief in body (1.4.2), and the philosophical beliefs in material (1.4.3) and immaterial (1.4.5) substances (SJ, chapters 5 and 6). Hume regards these beliefs as unjustified. They lack justification, according to Loeb, because they result from the propensity to ascribe identity to related objects (call it the *identity propensity*), and this propensity is unstable. It is unstable because it has unstable outputs—namely, the above-mentioned philosophical beliefs—in sufficient proportion to make it so (SJ 161).

On this interpretation, one would expect Hume to make the case that the identity propensity is unjustifying by showing that each philosophical belief results from the propensity and then arguing that each such belief is unstable.⁹ Is this what we find in the text?

Let us take as our first example of a philosophical belief the ancient belief in material substance. At T 1.4.3.1 (SBN219) Hume calls the ancient beliefs in material substance, accidents, occult qualities, and the like “capricious.” At T 1.3.9.6 (SBN 109) Hume explicitly ties caprice to unstable ideas, and to associations by resemblance, of which the identity propensity is a special case. Putting T 1.3.9.6 (SBN 109) together with T 1.4.3.1 (SBN 219), Hume is in a position to infer from the fact that these ancient beliefs are capricious that they are unstable. Moreover, Hume clearly ascribes the ancient belief in material substance to the identity propensity (T 1.4.3.4–5; SBN 220–1), as Loeb observes (SJ 148–9). I note, however, that the passage in which Hume makes this ascription is devoted to a purely psychological description of the origin of the ancient belief in material substance, with the exception of the veritistic term “deceives” at T 1.4.3.3 (SBN 220). Here there is no discussion of whether the ancient belief is justified. And when Hume comes to make the case that the ancient belief in material substance is unjustified, he does not argue that it is unstable or that it results from the identity propensity.

To be sure, there is a momentary link between the unjustifiedness of the ancient beliefs and their instability when Hume speaks of the ancient beliefs as “unreasonable and capricious” (T 1.4.3.1; SBN 219). It is safe to say that (as at T 1.3.9.6; SBN 109) “unreasonable” is a term of epistemic opprobrium meaning something like “unjustified,” and not merely a causal term denying that an idea arises from reason. Despite this momentary link between the unreasonableness and capriciousness of the ancient beliefs, Hume develops his case for the unreasonableness of the ancient beliefs without overt appeal to capriciousness.

Hume's case for the unreasonableness of the belief in material substance has two parts. He devotes most of his attention to arguing that the ancient belief in

accidents is unreasonable. He then infers from the unreasonableness of the belief in accidents that the belief in material substance is also unreasonable. Hume makes the case that the belief in accidents is unreasonable on the ground that it is inconsistent with the intuited premise “Every quality . . . may be conceiv’d to exist apart . . . from every other quality” (T 1.4.3.7; SBN 222). He then argues that the belief in material substance is unjustified on the ground that it entails the belief in accidents. The latter “is an unavoidable consequence of” the former (T 1.4.3.7; SBN 222). Neither his argument that the belief in accidents is unreasonable, nor his argument that the belief in material substance is unreasonable makes any obvious appeal to the instability of these beliefs. Falsity rather than instability may play the key role in Hume’s argument against the justifiedness of the belief in material accidents.¹⁰

Regarding Hume’s argument that the belief in material substance is unreasonable, we could interpret him as arguing that the belief is unjustified on the ground that its being justified would entail that any belief that follows from it is justified, and thus that the belief in accidents is justified; but the latter belief is unjustified, as we know by intuition. Alternatively, we could interpret Hume as arguing that the belief in material substance is unjustified on the ground that its justification is undermined by the recognition that it entails an unjustified or a false belief, the belief in accidents. On neither of these interpretations does his argument that the belief in material substance is unjustified appeal to instability. On both interpretations, the argument relies on garden-variety epistemic principles available to any viable account of justification.

Let us turn now to Hume’s argument against the justifiedness of the belief in accidents. Does this argument advert to instability? I can see only two ways to read the argument. On one reading, Hume shows that the belief in accidents is unjustified by showing that its denial is sanctioned by intuition. This suffices to show that the belief in accidents is unjustified because the sanctioning of the denial of the belief by intuition undermines the justification for the belief that would otherwise be provided by the identity propensity. On this reading, Hume’s argument depends on the principle that a belief is justified, other things equal, if it results from a mechanism of the right kind. It also depends on the principle that a belief is justified all things considered only if there is no available mechanism of a suitable sort that sanctions a contrary belief. In the case of the belief in accidents, the belief-forming mechanism, the identity propensity, could (as far as the argument goes) be the right kind, so that the belief is justified other things equal. But the justification of the belief in accidents is undermined by the availability of a suitable mechanism, intuition, that sanctions the denial of the belief. On this reading, Hume argues that the belief in accidents is unjustified all things considered (in light of the fact that there is available a mechanism that sanctions the denial of the belief), even though it is justified other things equal.

There is a second way to read Hume's argument. On this second way, he argues that the belief in accidents is unjustified because it is *false*, as shown by intuition. On this reading, Hume's argument makes a *partial* rather than a complete case for his conclusion. He does not argue in full that the belief in accidents is unjustified. Rather, he attempts a partial case against its justifiedness by exhibiting a fact that detracts from its justifiedness—the falsity of the belief. Presumably, falsity detracts from the justifiedness of the belief because it detracts from the justifying power of the identity propensity. On this second reading, Hume is not showing that the belief is unjustified all things considered, but rather making a partial case against its justifiedness other things equal. This reading is most plausible on a reliability interpretation. I believe that the tone of the text fits this reading better than the first, undermining reading.

These two readings are, as far as I can tell, the only options for Hume's argument against the justifiedness of the belief in accidents. Yet neither is compatible with the stability interpretation. The first, undermining reading is incompatible with Loeb's stability interpretation, as I understand it.¹¹ According to Loeb's interpretation, Hume makes a case against the other-things-equal justifying power of the identity propensity by showing that its output is frequently unstable. On this interpretation, Hume should criticize the identity propensity by exhibiting its unstable output. But on the undermining reading, Hume criticizes the justifying power of the identity propensity, in the case of the belief in accidents, only by saying that intuition undermines the justifiedness of the belief. This is a point about the availability of a countervailing mechanism. It does not entail that the ancient belief in accidents is unstable—at best that such a belief would be unstable in the belief system of a Humean philosopher. Thus, on the undermining reading, the average stability of the output of the propensity plays no role in Hume's critique of the justifying power of the propensity in the case of the belief in accidents, contrary to Loeb's interpretation. Moreover, the point that intuition undermines the justifiedness of the belief in accidents bears on the *all-things-considered* justifiedness of the belief, not on its other-things-equal justifiedness, as required by Loeb's interpretation. (From this it follows that the point does not bear at all on the justifiedness of the *vulgar belief in body*, since it does not bear on the other-things-equal stability or justifying power of the identity propensity. This is contrary to Loeb's interpretation, according to which the case against the justifiedness of the vulgar belief adverts in part to the instability of philosophical beliefs, including the ancient belief in material substance.) In short, the first reading is incompatible with Loeb's interpretation.¹²

The second, partialist reading of Hume's argument against the justifiedness of the belief in accidents is also incompatible with Loeb's interpretation, and indeed with any other stability interpretation. On this reading, Hume's argument does bear on the other-things-equal justifying power of the propensity, and on

the character of its output. But it bears only on the falsity of its output, not on its *stability*. Of course, the falsity of a belief may provide evidence of its instability, but I see no textual case that Hume adverts to falsity in support of instability here. Thus, neither reading of Hume's critique of the justifiedness of the ancient beliefs is compatible with a stability interpretation.

Let us turn now to Hume's argument against the justifiedness of the philosophical belief in body. Does a stability interpretation work for this philosophical belief? Let us first ask whether the belief is unstable. Hume does not remark that it is capricious, as he does for the belief in material substances. To reach the philosophical belief in body, we begin with the vulgar belief in body, which arises from the identity propensity in this way. The propensity leads us to suppose that interrupted but exactly resembling impressions are identical. Intuition leads us to recognize that these impressions differ. We resolve the contradiction between what the propensity tells us and what intuition tells us by supposing that these impressions continue to exist unperceived by us, so that there is an uninterrupted exact resemblance, satisfying both the propensity and intuition. The vulgar belief in body is simply this supposition of the continued existence of impressions unperceived. The philosophical belief in body emerges when we reflect that the vulgar belief is false, in light of the double vision experiment and observations of perceptual relativity. The philosopher attempts to honor this reflection by supposing a continued existence, not of unperceived impressions, but of something unperceived that is distinct from any perceptions. Hume charges that this new supposition is merely a "palliative" remedy (T 1.4.2.46; SBN 211). The philosophical belief is "liable to the same difficulties" as the vulgar belief, "and is over-and-above loaded with this absurdity, that it at once denies and establishes the vulgar supposition" (T 1.4.2.56, SBN 218; cf. T 1.4.2.46ff., SBN 211ff.). All this may suggest that the philosophical belief in body is unstable.

However, this suggestion is misleading. To be sure, Hume says that philosophers revert to the vulgar belief in body "upon leaving their closets" (T 1.4.2.53; SBN 216). The philosophical belief in body is unstable in one sense: it ceases when reflection ceases. But what is relevant to the question of the justifiedness of the philosophical belief in body, on both the actual and reflective stability accounts, is whether the identity propensity fixes the belief in philosophers *when they achieve their highest level of reflection*. The question, then, is whether the philosophical belief in body is unstable in the course of *philosophical reflection*. Thus, the fact that it ceases when reflection ceases does not show that it is unstable in the sense relevant to justification according to the stability account.

To judge whether the philosophical belief is stable, let us begin by noting that the vulgar belief in body is stable (as Loeb allows, 161). The contradiction from which the vulgar belief arises causes only a temporary instability, and it is not an instability in the *output* of the overall mechanism that yields the vulgar belief in

body. (On the stability account, what is relevant to justification is instability in the output of the mechanism, not states involved in the process that gives rise to the output.) Thus, when Hume says that the philosophical belief is “liable to the same difficulties” as the vulgar belief, the difficulties of contradiction and preliminary instability to which he refers do not impugn the justification of the philosophical belief on the stability account.

Does Hume, then, ascribe a relevant instability to the philosophical belief when he complains of the additional “absurdity, that it at once denies and establishes the vulgar supposition”? I don’t think so. He specifies the absurdity:

Philosophers deny our resembling perceptions to be identically the same, and uninterrupted; and yet have so great a propensity to believe them such, that they arbitrarily invent a new set of perceptions, to which they attribute these qualities. I say a new set of perceptions: For we may well suppose in general, but ‘tis impossible for us distinctly to conceive, objects to be in their nature any thing but exactly the same with perceptions. What then can we look for from this confusion of groundless and extraordinary opinions but error and falshood? And how can we justify to ourselves any belief we repose in them? (T 1.4.2.56; SBN 218)

I see no explicit worry about stability here, or in the longer discussion that precedes it at T 1.4.2.46ff. (SBN 211ff.).¹³ Granted, our inability to conceive philosophical body distinctly *could* destabilize the philosophical belief. But I see no textual evidence that Hume means to say that it does have such an effect. I do not find clear textual evidence that Hume takes the philosophical belief in body to be unstable, or uses its instability as a point against its justifiedness or against the justifying power of the identity propensity.

The explicit worry at T 1.4.2.56 (SBN 218) is not about instability. It is about groundlessness and error. Though I cannot go into the matter here, error comes to the fore in 1.4.4. Quite possibly, the point of that section is to show that if we follow reason and causal inference, we arrive at the conclusion that the philosophical belief in matter is false. Since causal inference is “the ultimate judge of all systems of philosophy” (T 1.4.4.1; SBN 225), Hume appears to be making an observation in 1.4.4 that bears unfavorably on the reliability of the identity propensity. To the extent that a philosopher will hold that bodies are made of matter, the observation tells *ad hominem* against the truth of the philosophical belief in body and further disfavors the reliability of the identity propensity.

In sum, I am uncertain that in Hume’s sampling of the philosophical beliefs that result from the identity propensity—in particular, the ancient belief in material substance and the vulgar and philosophical beliefs in body—there are enough unstable beliefs to make the propensity unjustifying according to the stability

account. The vulgar belief in body is stable, and the philosophical belief in body may well be stable in reflection. The beliefs in substances are unstable, but in the case of the ancient belief in material substance, Hume does not trace its unjustifiedness to its instability.

There is an additional difficulty with Loeb's explanation of the unjustifiedness of the vulgar belief in body. The vulgar belief is supposed to be unjustified because the identity propensity has unstable outputs. Since the vulgar belief is itself stable, the unstable outputs must be philosophical beliefs. Loeb specifies that these unstable beliefs are the beliefs in material and immaterial substance, which he claims are unstable "even for an unreflective person" (161). Yet, on the actual stability account, a subject's belief is justified in virtue of the average stability of the outputs the mechanism would yield if used at the highest level of reflection actually achieved by the subject. The trouble is that the philosophical beliefs in substance would seem not to be unstable at the highest level of reflection achieved by vulgar subjects. For they are not even formed at that level of reflection; they require philosophical reflection, and this presumably demands a degree of reflection a bit higher than the highest level achieved by the vulgar.¹⁴ The instability of the philosophical beliefs in substance is thus irrelevant to the justification of the vulgar belief on the actual stability account of justified belief.¹⁵ This is a serious drawback of Loeb's explanation of the unjustifiedness of the vulgar belief in body.

What, then, makes the vulgar belief in body unjustified? Loeb proposes that Hume is driven to deny justification to the belief because it is false (SJ 195–6). But on the stability interpretation, the falsity of the vulgar belief does not *make* it unjustified. So Loeb's suggestion is that, given the falsity of the vulgar belief, "Hume deems it intolerable that the vulgar belief should be justified" (SJ 196). This explains why Hume is driven to trace the vulgar belief to the identity propensity. That propensity is a more likely candidate for an unstable mechanism than the other source of the vulgar belief, custom-and-galley. And Hume needs to trace the vulgar belief to an unstable mechanism in order to find the belief unjustified.

Loeb's falsity explanation of the unjustifiedness for Hume of the vulgar belief in body does not square with the stability interpretation of justified belief. More exactly, the falsity of the vulgar belief in body gives Hume no *rational* motivation to find the belief unjustified on a stability interpretation. To be sure, the considerations that Hume raises to show that the vulgar belief is false would destabilize the belief if they were taken into account by the vulgar subject. But although these considerations occur to one with only "a very little reflection" (T 1.4.2.44; SBN 210), it is clear that they do not occur to unreflective people. If they did, the vulgar belief in body would be destabilized; yet it is not. Since the recognition that the vulgar belief in body is false does not destabilize the belief, the falsity of the belief does not disfavor its being justified, on the stability account.

On the reliability interpretation, by contrast, the falsity of the vulgar belief rationally favors the conclusion that it is unjustified. Hume could reason in this way: vulgar beliefs in body are false; to the extent that they make up a proportion of the output of custom-and-galley (for those beliefs arising from coherence) and the identity propensity (for those beliefs arising from constancy), they degrade the reliability of the propensities and reduce their justifying power. In the case of the identity propensity, there are false philosophical outputs as well, further degrading the justifying power of the propensity. This explains why for Hume vulgar beliefs from the identity propensity are less justified than vulgar beliefs from custom-and-galley.

5. Stability and the Evaluation of Justifying Mechanisms

Let me turn now to Loeb's treatment of the prominent passage at T 1.4.4.1 (SBN 225–6) in which Hume contrasts causal inference and the identity propensity with respect to their power to justify:

But here it may be objected, that the imagination, according to my own confession, being the ultimate judge of all systems of philosophy, I am unjust in blaming the antient philosophers for making use of that faculty, and allowing themselves to be entirely guided by it in their reasonings. In order to justify myself, I must distinguish in the imagination betwixt the principles which are permanent, irresistible, and universal; such as the customary transition from causes to effects, and from effects to causes: And the principles, which are changeable, weak, and irregular. . . . The latter are neither unavoidable to mankind, nor necessary, or so much as useful in the conduct of life; but on the contrary are observ'd to take place only in weak minds, and being opposite to the other principles of custom and reasoning, may easily be subverted by a due contrast and opposition.

Loeb interprets this passage and the following paragraphs as explaining why causal inference is justifying and the identity propensity is not, by appeal to a difference in stability (SJ 154–62). But there is an alternative to this interpretation.

The passage begins with an objection. The objection is that Hume is “unjust in blaming the ancient philosophers for making use of” an imaginative mechanism (the identity propensity) to form beliefs because he himself uses an imaginative mechanism (causal inference) as “the ultimate judge of all systems of philosophy.” On the simplest reading of this opening objection, Hume is charged with being unjustified (“unjust”) in his evaluation (“blaming”) of the identity propensity as unjustifying if his evaluation of the propensity is true—if, that is, the propensity really is unjustified, as he claims. For he arrives at this evaluation using another

imaginative mechanism (causal inference). He must do so, since all philosophical systems (and hence the mechanisms that give rise to such systems) are evaluated using imagination (causal inference). But imaginative mechanisms are in general no more justifying than the identity propensity. So Hume's evaluation is no more justified than the mechanism it criticizes. The upshot of this objection to Hume's evaluation is that the evaluation is itself unjustified, if it is true.

What I want to call attention to is that, if this simple reading of the opening objection is correct, Hume avers at T 1.4.4.1 (SBN 225) that he uses causal inference to evaluate the identity propensity as unjustifying. This averral imposes a key constraint on an interpretation of justified belief: any interpretation must understand justification in such a way that the evaluation of the identity propensity as unjustifying depends on causal inference. Indeed, Hume's superlative, "*the ultimate judge of all philosophical systems*" (my italics), strongly suggests that evaluations of mechanisms that produce philosophical systems depend *centrally* on causal inference. My question, then, is whether the stability interpretation meets this key constraint. Does the evaluation of the identity propensity as unjustifying depend centrally on causal inference given the stability interpretation of "unjustifying"?

To simplify the discussion, I will assume at first an "averaging" version of the stability account: a propensity is stable just in case it *actually* yields stable ideas *on average*. I take it that we could infer the judgment that the identity propensity is unstable in this way:

The identity propensity yields just the particular ideas I1, I2, . . . , In.

Most of the ideas I1, I2, . . . , In are unstable.

So, by arithmetic, the identity propensity yields ideas that are unstable on average.

So, by the meaning of "unstable," the identity propensity is unstable.

Does this inference depend on causal inference? It does, but in my opinion it does not depend as centrally on causal inference as the passage suggests the evaluation of the justifying power of a mechanism must. The two steps in the inference itself are *deductive*. So the conclusion does not depend on causal inference in virtue of the character of these inferential steps; it must depend on causal inference for the premises. Clearly the first premise does derive from causal inference. But the way it does so seems not to be central enough to match Hume's superlative in the passage, "*the ultimate judge of all systems of philosophy.*" We would not call an inferential mechanism the ultimate judge of a system if it contributed to judging the system only by judging which output results from the mechanism used in the system. The gravity of Hume's language suggests that he has in mind something more central to our evaluation of the mechanism than merely which ideas it forms—namely, whether the mechanism has the justification-making property (i.e., stability, on the stability account). On the averaging stability account, this property is the aver-

age stability of the mechanism's actual output ideas. Thus, our question comes to whether causal inference judges that the mechanism's output is, or is not, stable on average. Equivalently, it comes to whether the second premise of the inference (that most of the ideas I_1, \dots, I_n are unstable) derives from causal inference.

To judge this matter, consider the judgment that idea I_1 is unstable. Assume for the moment a *conditional regularity* view of ideas, on which an idea is a conditional regularity according to which the will exhibits certain features given certain mental circumstances. Assume also the simple view that an idea is stable during a period of time just in case the relevant conditional regularity obtains through sufficiently much of that period. Then an idea is unstable during a period just in case the conditional regularity fails to obtain often enough in the period. Given this definition, to judge that an idea is unstable, we judge that the conditional regularity fails to obtain often enough. Judging that the conditional regularity fails to obtain is a matter of first observing that in a high enough percentage of sampled cases, the triggering mental circumstances obtain but the will fails to exhibit the required features, and then inferring the absence of a conditional regularity. Clearly, this requires only an enumerative induction from the absence of sampled conjunctions to the absence of the conditional regularity. From this, we infer by deduction (in the presence of the definition of "unstable") that the idea is unstable. This reasoning involves no inference from effect to cause (or vice versa, or even from the absence of an effect to the absence of a cause). Thus, it involves no causal inference. On the conditional regularity view of ideas, then, we may judge whether an idea is unstable without employing causal inference. And thus we can derive the second premise of the inference that the identity propensity is unstable without employing causal inference.¹⁶

Is this conclusion upset when we swap the conditional regularity view of ideas for a more sophisticated *causal* view of ideas? Suppose an idea is a conditional *causal* regularity according to which the will is caused to exhibit certain features given certain mental circumstances. Must one employ causal inference to judge whether an idea is unstable on this account of ideas? Granted, one must judge that there is a causal regularity to judge that the relevant conditional causal regularity is satisfied over time—and this judgment would require inference from an effect to a cause. But to judge that the conditional causal regularity is *not* satisfied in an interval of time, one need only observe that willings often fail to obtain when the relevant mental circumstances obtain, infer by enumerative induction that a conditional regularity fails to obtain, and then infer by deduction that the conditional causal regularity fails to obtain. This involves no nondeductive inference from effect to cause, or from the absence of an effect to the absence of a cause. Thus, it involves no causal inference. So it is possible to judge that an idea is unstable without employing causal inference, even on the causal dispositional view of ideas. Of course, one could also judge that willings that in fact often obtain under the relevant mental

circumstances are often not *caused* by these circumstances. This judgment *would* depend on a causal inference. But this point does not undermine my claim that one could judge that willings often fail to obtain when the relevant circumstances obtain without employing a causal inference.¹⁷ On both the conditional regularity and the conditional causal regularity views of ideas, then, the judgment that the identity propensity is unstable need not derive from causal inference.¹⁸ If so, on the averaging stability account we have so far assumed, judging that a mechanism is unjustified need not involve causal inference, contrary to Hume's averral at T 1.4.4.1 (SBN 225). This tells against the averaging stability account.

One might object that a causal inference to the instability of the identity propensity is needed if we replace the averaging stability account with a *tendency* stability account on which a mechanism's stability is its tendency to yield stable ideas. For inferring from the instability of actual and counterfactual ideas to the propensity's tendency to produce unstable ideas requires an inference from effect to cause. Thus, one might object, a tendency stability account is consistent with Hume's averral of causal inference at T 1.4.4.1 (SBN 225).

But first, even if this is right, most of the evaluative work is still being done by the judgments of the instability of particular ideas, so it is not clear that this is the grand use of causal inference suggested by the passage. Second, I concede that causal inference is needed to judge that these ideas are (actual or counterfactual) *outputs* of the propensity. But I take it that once the actual and counterfactual outputs of the propensity are identified, judging its tendency to produce unstable ideas is a matter of checking whether these outputs are sufficiently representative to infer its average stability in the totality of its outputs. And this requires only an enumerative induction from the observed ratio of instability to the total ratio. No *further* causal inference is needed. At any rate, a tendency stability account on which the judgment of instability is made in this way would seem to be as defensible as any on which one needs a further causal inference. I conclude that there is nothing in a stability account per se, even a tendency account, which requires a grand use of causal inference. For this reason, if Hume held a stability account, it would be open to him to respond to the objection at T 1.4.4.1 (SBN 225) by evasion rather than in the way he actually responds: he could simply answer the objection by retreating to a well-motivated stability account that requires no grand use of causal inference, rather than admitting that causal inference is required and arguing that it is not relevantly like the identity propensity.

The point I wish to make against the stability interpretation, then, is this. *Treatise* 1.4.4.1 (SBN 225) suggests that the evaluation of the identity propensity depends in a certain way on causal inference. But the stability interpretation entails that the evaluation of the propensity need not depend on causal inference in this way: the inference need only involve enumerative induction and deduction from observations. So the passage counts against the stability interpretation.

I conclude that the stability interpretation does not square with the simple reading of the objection at T 1.4.4.1 (SBN 225). The reliability interpretation fits the objection better. On the reliability interpretation, to evaluate the identity propensity as unjustified, Hume judges that it is unreliable. To make the latter judgment, Hume judges that the vulgar belief in body, the philosophical belief in material substance, and (in my view) the philosophical beliefs in body and matter are false, and he infers the unreliability of the identity propensity. Hume makes the case that the vulgar belief in body is false by an experiment (double vision) that depends on causal inference. And he judges the philosophical beliefs in body and matter false by causally inferring that primary qualities are not distinct from perceptions. In these cases, the use of causal inference is as obvious and central to the evaluation of the propensity as it could be. I conjecture that Hume refers precisely to these causal inferences when he avers at T 1.4.4.1 (SBN 225) that imagination is the ultimate judge of philosophical systems.

We have focused so far on the opening objection at T 1.4.4.1 (SBN 225). Let us turn now to the different question whether Hume's *reply* to this objection supports the stability interpretation. On the simple reading of the objection I elaborated above, the charge is one of *epistemic inconsistency*: if Hume's evaluation of the identity propensity as unjustified is *true*, then this evaluation is itself *unjustified*. For this evaluation derives from causal inference. But causal inference is similar to the identity propensity in respects relevant to whether the latter mechanism justifies the evaluation of the propensity as unjustifying. Hence causal inference does not justify that evaluation if the identity propensity is not a justifying mechanism. Thus, Hume's negative evaluation of the identity propensity is either *false* or *unjustified*.

On the most straightforward reading of Hume's reply to this objection, he answers it by conceding that causal inference is an imaginative mechanism while denying that causal inference is like the identity propensity in respects relevant to whether it justifies his evaluation of the propensity as unjustifying.¹⁹ Causal inference is more justifying than the identity propensity is. Hence, causal inference can justify the evaluation of the propensity as unjustifying, despite the fact that both are imaginative mechanisms and the identity propensity is unjustifying. Loeb takes the crucial difference between the mechanisms to be that the identity propensity is "subverted by a due contrast and opposition" with causal inference. Loeb then reads subversion as meaning that causal inference more frequently destabilizes the outputs of the identity propensity than conversely. Presumably, this claim of asymmetrical destabilization partially supports the claim that causal inference is more justifying (at the level of reflection involved in the evaluation of the identity propensity) than the identity propensity is, assuming the actual stability account of justification. For the claim shows that causal inference is, other things equal, more stable than the identity propensity is (at this level of reflection)—just what

is required for a difference in justifying power, on the actual stability account. Thus causal inference can justify Hume's evaluation of the identity propensity as unjustifying even if that evaluation is true.

I have two points to make against this asymmetrical destabilization reading of subversion. My first point is a conditional one: that the asymmetrical destabilization account of subversion is at odds with the actual stability account of justification *if* it is intended as a general account of the relative justifying power of mechanisms for all beliefs. On the asymmetrical destabilization account, causal inference is more justifying than the identity propensity is because under "a due contrast" it more often destabilizes the output of the propensity than vice versa. I am not now raising a problem for the account applied to the justifying power of mechanisms in producing the beliefs of reflective people, but the account is incompatible with the actual stability account when applied to the beliefs of *unreflective* people. So applied, the asymmetrical destabilization account must make sense of a due contrast between causal inference and the identity propensity for the beliefs of unreflective people, including the vulgar belief in body. Causal inference must, under a due contrast, destabilize the vulgar belief in body. But there is no reading of "a due contrast" compatible with the actual stability account of justification. Hume cannot mean that causal inference *actually* destabilizes the vulgar belief, since it clearly does not do so. Could he have in mind that causal inference *would* destabilize this output in unreflective people *were they to reflect*? But the fact that causal inference would destabilize the vulgar belief in unreflective people were they to reflect is irrelevant to the justifying power of causal inference on the actual stability account. On that account, the justifying power of causal inference is a function of the stability of the judgments that it actually makes—in particular, the judgment in reflection that body in the vulgar sense does not exist. The judgment that body in the vulgar sense does not exist is justified because causal inference is stable in reflection. On the actual stability account, this justification-making property is *not* that causal inference would destabilize the output of the identity propensity in *unreflective* people *were they to reflect*. Rather, it is that causal inference *actually* destabilizes the output of the identity propensity *in reflective people*. Thus, in the case of the vulgar belief in body, there is no room for destabilizing under a due contrast in the sense required by the actual stability account. To put it differently, if a due contrast involves *actual* reflection, causal inference does destabilize some outputs of the identity propensity (the vulgar and philosophical beliefs in body) and not vice versa; but this does not bear on the justifying power of the mechanisms in the case of unreflective beliefs in body, according to the actual stability account. If a due contrast involves *counterfactual* reflection, again the power of causal inference to destabilize the vulgar belief in body does not bear on the justification of unreflective beliefs according to the actual stability account. The only sense

in which causal inference destabilizes the vulgar belief in body in unreflective people under a due contrast is a sense irrelevant to the justifying power of causal inference in any of its outputs, on the actual stability account of justification. Thus, the actual stability account is at odds with the asymmetrical destabilization reading of subversion if the latter is intended as a general account of justification. Of course the context of T 1.4.4.1 (SBN 225) is a reflective evaluation of mechanisms and systems, and Hume might intend the subversion account to be restricted to reflective beliefs. So my conditional objection to the asymmetrical destabilization reading is hardly compelling.

My second point against Loeb's reading of subversion as destabilization is, I think, more troubling: given the context, the contrast to which Hume refers in speaking of subversion is plausibly taken to be, most immediately, a contrast between the use of causal inference to evaluate whether the identity propensity is justifying, and the use of the identity propensity to evaluate whether that same propensity, the identity propensity, is justifying. For what is at stake in the discussion is whether causal inference is epistemically preferable to the identity propensity when it evaluates the propensity as unjustifying. So causal inference and the identity propensity are supposed to contrast in their evaluations of the identity propensity. But this rules out the view that an evaluation of the identity propensity is a judgment of its *stability*. For causal inference and the identity propensity do not contrast in their judgments of the stability of the propensity: unlike causal inference, the propensity does not even judge its own stability. In other words, the fact that the two mechanisms contrast in their evaluations of the propensity rules out the stability account of justification. I observe, however, that the mechanisms do contrast in their evaluations of the propensity if these evaluations are judgments of the *reliability* of the propensity. For they contrast in their judgments of the truth values of the vulgar belief in body and perhaps of the philosophical belief in body. And so they contrast in their judgments of the reliability of the identity propensity. Thus, the best interpretation of the contrast of which Hume speaks at T 1.4.4.1 (SBN 225) depends on a reliability interpretation of justification. The best understanding of Hume's talk of being "subverted by a due contrast" seems to be veritistic. Hume's reply to the objection at T 1.4.4.1 (SBN 225), as well as the objection itself, seems to call for a veritistic interpretation.

In this article I have pointed to some passages in Book 1 of the *Treatise* where a veritistic alternative to the stability interpretation seems to explain the text at least as well as the stability interpretation does. To say this is not to say that any alternative interpretation fares as well overall as the stability interpretation. Loeb is right that there is a good deal more talk of stability in Hume than of truth and reliability. Indeed, Loeb has done his work so well that as things now stand, the stability interpretation is the best we've got. Loeb's book has taken this reader closer to the heart of Hume's epistemology than any other work on Hume.

NOTES

I would like to thank Kevin Kimble and Ken Winkler for help on this article.

1 (Oxford: Oxford University Press, 2002).

2 Quotations from Hume's *A Treatise of Human Nature* ("T") are from the edition prepared by David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2000). I refer to passages by book, part, section, and paragraph, followed by a page reference to David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd ed., revised by P. H. Nidditch ("SBN") (Oxford: Clarendon Press, 1978).

3 However, Loeb does nod in the direction of a broader account, to include adaptive-ness, at 125–6 n36.

4 Loeb argues persuasively that second-order general rules—generalizations about the success of classes of first-order generalizations—are stabilizing for Hume (SJ 105ff.). But these are not associated with second-order dispositions that are fully general for belief. The dispositions associated with second-order general rules would seem to be present at best only for beliefs that result from these generalizations. And these dispositions are not general across the population: "wise men" rather than the vulgar "are commonly guided by" the rules (T 1.3.13.12; SBN 150).

5 The alternative to an epistemic reading of "founded on so little reason" is to read the phrase as denying that such fictions result from causal inference. But Hume cannot merely be saying that the fictions produced by associations of resemblance and contiguity are not produced by causal inference. From such a meager premise it would hardly follow, as Hume takes it to follow, that these fictions result from will. Moreover, assuming that these fictions are not produced by causal inference would *presume* that they are not for that reason beliefs. But Hume is precisely trying to *show* that these fictions are not beliefs.

6 I offered a sketchy reliability interpretation of Hume's account of justified belief, with attention to 1.4.2, in *Knowledge and Belief* (London: Routledge, 1992), chapter 3.

7 It is also true on the alternative plausible requirements that the mechanism yield stable beliefs over the portion of the subject's system—reflective or unreflective—to which the given belief belongs, or over the belief systems of all subjects.

8 I side with Loeb against Michael Williams in thinking that Hume does not take a contextualist approach, assigning a different justificatory status to the causal inferences of a reflective person during everyday life and during reflection (SJ 98–100). In 1.4.7, Hume concludes that causal inference is unjustifying for a reflective person, not only in the study, but during everyday life. But unlike Loeb, I take this to tell against a stability interpretation. On a well-motivated stability account, a reflective person's everyday beliefs from causal inference ought to be justified because causal inference is stable in the everyday belief system of a reflective person even if it is unstable in the study.

9 I note that on the stability interpretation the identity propensity could be challenged on the ground that it is relevantly similar to (or cannot be clearly distinguished from) such obviously unstable mechanisms as the propensity to add a quality to a resembling

object. But this challenge would not establish firmly that the identity propensity is unstable and so would not yield Hume's firm result that the beliefs in body and substance are unjustified.

10 Granted, there is another mood in which Hume castigates not the truth but the intelligibility of the proposition that there are substances (e.g., T 1.4.5.6; SBN 234). In this mood, Hume denies that it is "reasonable . . . to enter seriously into the dispute" (T 1.4.5.2; SBN 232): epistemic epithets do not apply. Loeb distinguishes considerations of meaning from epistemic considerations on 172.

11 I find the undermining interpretation implausible. First, there is the point that at T 1.4.4.1 (SBN 225) Hume says that causal inference judges all philosophical systems, by which he means it judges the justifying power of the mechanisms that give rise to them. It is plausible that at T 1.4.3.2 (SBN 219) intuition is playing a like role in judging the justifying power of the identity propensity, as it gives rise to the belief in accidents. If so, the role of intuition is not to undermine the propensity but to judge it—by judging its output false. Second, the use of intuition to judge the ancient belief in accidents false parallels Hume's earlier use of causal inference to judge the vulgar belief in body false. And in this case, Hume cannot have in mind the undermining of the identity propensity by causal inference, since the relevant causal inference is not available to the vulgar.

Hume does question whether the Peripatetics who believe in "*sympathies, antipathies, and horrors of a vacuum*" have any "excuse" for their reliance on "every trivial propensity of the imagination" (T 1.4.3.11; SBN 224–5). It may be that this question is intended to insinuate the availability of undermining reflection in the case of these particular Peripatetic beliefs. But these beliefs derive from the inclination to bestow emotion on external objects, not from the identity propensity. This inclination is perhaps an instance of the propensity to add a quality to a resembling object.

12 To be sure, the first reading is compatible with ascribing a different, comparative version of the stability interpretation on which intuition undermines the justifiedness that the identity propensity would otherwise supply the belief in accidents because intuition is available (to a sufficiently reflective subject) and more stable than the identity propensity. But this comparative stability interpretation has no rationale in the value of actual stability, and there is little evidence that this is what Hume has in mind at T 1.4.3.1–2; SBN 219.

13 Loeb does concede that the philosophical beliefs are more stable "than the contradictory beliefs they obscure" (SJ 172).

14 I note, too, that an amended actual stability account on which the stability of a mechanism in a given case is the average stability of its outputs in all cases runs into trouble here. If we count each actual belief equally, then on this amended account, vulgar beliefs in body come out justified, since they are stable beliefs far more numerous than the unstable philosophical beliefs produced by the identity propensity. But if we count each *type* of belief equally (where type is defined by propositional content), then causal inference comes out unjustifying, since its unstable reflective outputs will be as numerous as its unreflective outputs.

15 There is a related worry about Loeb's account of the unjustifiedness of the vulgar belief. Perhaps it is the palliative character of the remedy in the philosophical use of the

identity propensity that deprives the propensity of its power to justify the philosophical beliefs. This character does not, however, carry over to the use of the propensity to form the vulgar belief in body. The worry, then, is that the instability in the philosophical beliefs does not detract from the propensity's power to justify the vulgar belief.

16 I take it that the same goes for a dispositional view of ideas, since a disposition is not (or at any rate need not be viewed as) a cause of the observed conjunction.

17 Could Hume infer the instability of the identity propensity in a different way? One proposal would be that he abstracts a description of the propensity from descriptions of the causes of various beliefs of interest (the beliefs in body and substrata). He then examines the intrinsic character of the propensity so described to judge whether it would yield unstable beliefs. It is not entirely clear that this would require a causal inference from the description to the conclusion that the beliefs are unstable. One might be able to deduce the instability of the propensity from the description. But Hume does not in fact, and cannot, judge the identity propensity unstable in this way. For the intrinsic character of the propensity varies from one use to another: it does not make the vulgar belief in body unstable. Since the propensity lacks a uniform intrinsic character involving instability, Hume could not infer its being unstable from a description of it, either by causal inference or by deduction.

18 I grant that, on the causal view of ideas, a *positive* judgment that a mechanism is stable *does* depend grandly on causal inference, in the way the passage suggests. But this does not change the fact that a judgment that a mechanism is *unstable* need not depend on causal inference in this way. I note too that causal inference is for Hume "the ultimate judge of *all* systems of philosophy" (my emphasis), not just of the systems he evaluates positively. Thus, causal inference is the ultimate judge of the negatively evaluated philosophical beliefs in body and substrata.

19 I assume the straightforward reading for purposes of discussion, but I doubt that the respects Hume cites are supposed to be relevant to whether causal inference is *justifying*. Hume replies to the charge of epistemic circularity by adverting to the permanence, irresistibility, universality, and subversive power of causal inference. Hume's reply must answer the charge that he has in a certain sense begged the question in using causal inference to conclude that a relevantly similar mechanism, the identity propensity, is unjustifying. His reply is that causal inference differs from the identity propensity in being permanent, etc. Yet it takes causal inference to judge that causal inference differs in this regard. Would Hume reply to the charge of begging the question by using causal inference to justify the conclusion that causal inference is justifying? I think not. I grant that he does once use causal inference to justify the conclusion that causal inference is justifying (T 1.3.15.6; SBN 173–4). But he does this in a context in which the justifying power of causal inference has not yet been questioned and in which he proceeds on the pre-reflective intuition that causal inference is justifying. I also grant that at T 1.4.4.1 (SBN 225) he uses causal inference to justify the conclusion that causal inference has an important normative status. But I doubt that this status is being *justifying*. I suspect it is being the proper method to make philosophical judgments as to which beliefs are justified. On the best reading of the passage, Hume is looking for a method that can close the interminable debate over philosophical issues like material substances. For this purpose, nothing is more promising than a method that *everyone* uses and will forever *continue* to use, the output of which cannot

be *resisted*—a mechanism “permanent, irresistible, and universal.” Hume does use causal inference to reach his conclusion that causal inference has the desired features. And this does justify his conclusion because causal inference is in fact justifying. But all that Hume attempts here is to reach the conclusion that causal inference has the right features to afford a proper philosophical judgment of which mechanisms are justifying, by using the proper philosophical method. Of course, this effort too is question-begging, but at least it begs a different question from the one Hume sets out to answer. I grant that closing debate entails stability, but this aim defines philosophical inquiry and not cognition that influences the will, to which the status of justification applies. Permanence, irresistibility, and universality do not obviously matter in the latter case. I observe, finally, that whether we endorse the straightforward reading of Hume’s reply, or instead adopt the reading I propose in this note, does not much affect my objection to the stability interpretation of T 1.4.4.1 (SBN 225): on either reading, at least my second point against the stability interpretation of Hume’s reply applies, as does my point against the stability interpretation of the objection Hume addresses at T 1.4.4.1 (SBN 225).

