

Hume's Vicious Regress

1 Promissory Note

Here are three central questions in Hume scholarship.¹ First, how does his psychological naturalism relate to his apparent avowals of causal subjectivism? Second, what contradiction is he wrestling with in the Appendix to the *Treatise*? Third, did he change his mind about the objectivity of causation between the *Treatise* and the first *Enquiry*? I'll answer these questions by tracking Hume's thinking on causation and personal identity from the publication of Books 1 and 2 of the *Treatise* in 1739 to the publication of the first *Enquiry* in 1748.

I aim to show that, in Book 1 of the *Treatise*, Hume argues that every causal relation (ordinarily so-called) between external objects depends on an actual and distinct association of ideas. Later in the Book, he generalizes this argument and argues that every seemingly real bond between objects, including the relation that ties a person's perceptions together, ought to be analyzed as depending on distinct associations of ideas. In the Appendix, he realizes that his mania for analyzing relations through associations of ideas leads to an infinite regress. In the first *Enquiry*, he abandons the position that causal relations depend on the existence of some distinct association in some actual observer. This avoids the regress and reduces the subjectivism in Hume's account of causation to a mere echo.

¹ I presented part of this paper at the 2005 South Central Seminar in the History of Early Modern Philosophy, and I thank the participants for their help. I also thank Martin Curd and anonymous reviewers for comments on drafts and Alex Rajczi for comments on multiple drafts.

2 Associations of Ideas in the *Treatise's* Account of Causation

2.1 Hume's Answer to Malebranche's Question

In this section, I present a *prima facie* case that Hume's account of causation in *Treatise* 1.3.14 makes the relation depend on the inferences of observers and that we should take this thesis as an earnest piece of metaphysics. He begins the section ('Of the idea of necessary connexion') with his account of the origin of our idea of necessity or efficacy. We perceive pairs of objects that are contiguous in time and place. After we perceive enough resembling pairs a habit arises in us so that, with the appearance of one object, the mind is determined to form a lively idea of the other. This determination gives us the impression of necessity from which the idea of necessary connection is copied.

Hume expects that these psychological "consequences will at first sight be receiv'd without difficulty, as being evident deductions from principles, which we have already establish'd, and which we have often employ'd in our reasonings" (*THN* 1.3.14.2). He cautions his readers to show the degree of caution appropriate for a world-shaking result, since "I have now examin'd one of the most sublime questions in philosophy, *viz. that concerning the power and efficacy of causes*; where all the sciences seem so much interested" (*ibid.*). What he means by 'the efficacy of causes' is "that quality which makes them be follow'd by their effects" (*THN* 1.3.14.3).

Let us linger for a moment on this rhetorical turn. Hume assumes that his readers will accept his account of the origin of our idea of necessity complacently, but he feels obligated to slow things down. It's "proper to give warning" (*THN* 1.3.14.2) that his story implicitly solves a contentious problem. He plainly takes his account of how causes bring

about effects to be more impressive than his psychological description of the origin of our idea of necessity.

We might think that the question is empty rather than sublime and that there is no quality common to all causes through which they bring about their effects. According to Hume, however, it is one of the central problems in the history of philosophy: “there is no question, which on account of its importance, as well as difficulty, has caus’d more disputes both among antient and modern philosophers” (*THN* 1.3.14.3). His polemical remarks against rival accounts give us an idea of what he has in mind. On some scholastic accounts, the efficacy of causes lies in “substantial forms, and accidents, and faculties” (*THN* 1.3.14.7); on Cartesian accounts, the efficacy that produces motion “must lie in the DEITY, or that divine being, who contains in his nature all excellency and perfection” (*THN* 1.3.14.9). These aren’t pedestrian examples of causes, but rather foundational entities that undergird general theories of causation.

The reference to the Cartesians is especially illuminating, since Malebranche’s *Search After Truth* was the first item on a list of background reading for the *Treatise* that Hume offered his friend Michael Ramsey.² At the beginning of Hume’s criticisms of “those philosophers, who have pretended to explain the secret force and energy of causes” by appealing to something in bodies, he drops a footnote by way of acknowledgment and endorsement to Book 6, Part 2, Chapter 3 of the *Search* and to the Elucidation of that chapter (*THN* 1.3.14.7n29). Indeed, as Charles McCracken writes, “there are places in his discussion of causality that betray the influence not merely of Malebranche’s ideas about

² Charles J. McCracken, *Malebranche and British Philosophy* (Oxford: Clarendon Press 1983), 254.

causality, but even his words.”³ I say that Hume’s question of the efficacy of causes is Malebranche’s question of the efficacy behind what are usually called causes.

Malebranche’s answer to that question is God. In contrast, Hume concludes,

The efficacy or energy of causes is neither plac’d in the causes themselves, nor in the deity, nor in the concurrence of these two principles; but belongs entirely to the soul, which considers the union of two or more objects in all past instances. ‘Tis here that the real power of causes is plac’d, along with their connexion and necessity (*THN* 1.3.14.23).

The three theories that Hume treats as rivals correspond to the three possibilities that Malebranche considers in the Fifteenth Elucidation: pagan naturalism, occasionalism, and concurrentism. Hume thinks of the soul as a fourth answer to Malebranche’s question.

The more seriously we take these signs that Hume is pursuing Malebranche’s metaphysical question, the more seriously we’ll take the evidence that Hume is committed to a robust subjectivism in causation. One piece of evidence that Hume takes causal connections to depend on causal inferences is the following piece of foreshadowing:

having found, that after the discovery of the constant conjunction of any objects, we always draw an inference from one object to another, we shall now examine the nature of that inference, and of the transition from the impression to the idea. Perhaps ‘twill appear in the end, that the necessary connexion depends on the inference, instead of the inference’s depending on the necessary connexion (*THN* 1.3.6.3).

Less hesitantly, after he investigates the psychological questions surrounding causal inference, Hume returns to the question of the nature of causation with the following apology:

This order wou’d not have been excusable, of first examining our inference from the relation before we had explain’d the relation itself, had it been possible to proceed in a different method. But as the nature of the relation depends so much on that of the inference, we have been oblig’d to advance

³ *ibid.*, 257.

in this seemingly preposterous⁴ manner, and make use of terms before we were able exactly to define them, or fix their meaning (*THN* 1.3.14.30).

For all the world it seems that Hume takes causal relations to depend on causal inferences, and these to rest on custom. It's as if associations of ideas are ur-causes upon which all the ordinary causes depend.

In 1.3.14, Hume has a shocked interlocutor summarize his views as follows: "What! the efficacy of causes lie in the determination of the mind! As if causes did not operate entirely independent of the mind, and wou'd not continue their operation, even tho' there was no mind existent to contemplate them, or reason concern them" (*THN* 1.3.14.26). Given the author of his voice, we know that the interlocutor hasn't misunderstood Hume's view.

Conversely, Hume attributes to his opponent the view that "thought may well depend on causes for its operation, but not causes on thought. That is to reverse the order of nature, and make that secondary, which is really primary" (*ibid.*). In the next paragraph, Hume construes causation on his opponent's view as a "real connexion" and as a "real intelligible connexion" (*THN* 1.3.14.27). Three points should be made about real connections. First, given the interlocutor's outcry, they must be relations that bind their relata together in an observer-independent way.⁵ Second, if the complaints that Hume considers against his definitions are motivated by the same assumption that causation is a real connection, then real connections do not depend on any "objects foreign" to the relata (*THN* 1.3.14.31). Third, Hume assumes that if we could perceive a real connection between objects, we could infer the presence of one from the presence of the other independently of

⁴ 'Preposterous' in the obsolete sense of *in an inverted order*, as David and Mary Norton observe in their editorial notes to the *Treatise* (467n30).

⁵ See Louis Loeb's discussion of causation as an intrinsic relation in Hume in 'Causation, Extrinsic Relations, and Hume's Second Thoughts about Personal Identity' [Relations], *Hume Studies* 18 (1992): 219-31, at 220-22.

experience.⁶ In discussing the inferences of animals he writes, “Beasts certainly never perceive any real connexion among objects. ‘Tis therefore by experience they infer one from another” (*THN* 1.3.16.8). The premise supports the conclusion only on the assumption the perception of real connections between objects would make an appeal to experience superfluous.

In order to move from his psychological account of the origin of our idea of efficacy to a rival to Malebranche’s account of efficacy, Hume needs a bridge principle. He appeals to his empiricist theory of representation and argues that studying the conditions under which the impression of necessity arises will illuminate the corresponding idea:

Ideas always represent their objects or impressions; and *vice versa*, there are some objects necessary to give rise to every idea. If we pretend, therefore, to have any just idea of this efficacy, we must produce some instance, wherein the efficacy is plainly discoverable to the mind, and its operations obvious to our consciousness or sensation. . . . Our present business, then, must be to find some natural production, where the operation and efficacy of a cause can be clearly conceiv’d and comprehended by the mind, without any danger of obscurity or mistake (*THN* 1.3.14.6).

Hume tries to track down the circumstances that give rise to the impression of necessity partly because he think that ideas represent the corresponding impression. But he also seems to assume that finding the conditions under which the impression of efficacy is produced will reveal a clear case of efficacy itself.

As Robert McRae observes,⁷ Hume follows Locke (E 3.3.10, 3.4.4) in believing that words that words that signify simple ideas are, in a sense, indefinable. Hume thinks that the best we can do in giving an account of the signification of simple impressions is to give an account of the causal nexus in which they occur. So, for example, since “the passions of

⁶ P.J.E. Kail calls this ‘the Bare Thought’ of causal power, *Projection and Realism in Hume’s Philosophy [Realism]* (Oxford: Oxford University Press, 2007), 84.

⁷ Robert McRae, ‘Hume on Meaning’ [‘Meaning’], *Dialogue* 8 (1969): 486-91, at 489.

PRIDE AND HUMILITY” are “simple and uniform impressions, ‘tis impossible we can ever, by a multitude of words, give a just definition of them, or indeed or any of the passions” (*THN* 2.1.2.1). The best we can do “is a description of them, by an enumeration of such circumstances, as attend them” (*ibid.*). Hume’s account of pride thus consists in giving us the circumstances that produce the impression of pride and describing the effects of that impression: “here then is a passion plac’d betwixt two ideas, of which the one produces, and the other is produc’d by it. The first idea, therefore, represents the *cause*, the second the *object* of the passion” (*THN* 2.1.2.4). Hume’s account of efficacy is like his account of pride: it explains the relevant idea by describing the causes and effects of the corresponding impression.⁸

Hume’s project is to find the source of the impression of the efficacy. Some scholastic accounts are so far from appealing to vivid impressions of efficacy that they appeal to principles which “are not in reality any of the known properties of bodies, but are perfectly unintelligible and inexplicable” (*THN* 1.3.14.7). Hume rejects Malebranche’s occasionalism because he rejects Malebranche’s theory of ideas:

the principle of innate ideas being allow’d to be false, it follows, that the supposition of a deity can serve us in no stead, in accounting for that idea of agency, which we search for in vain in all the objects, which are presented to our senses, or which we are internally conscious of in our own minds (*THN* 1.3.14.10).

By the 1730s, Lockeanism had carried the day, and hardly anyone in Britain or France thought that any of our ideas were immediate gifts from God.

The only thing that provides us with the requisite impression is an association of ideas: “there is no internal impression, which has any relation to the present business, but that propensity, which custom produces, to pass from an object to the idea of its usual

⁸ *ibid.*, 489-91.

attendant. This therefore is the essence of necessity” (*THN* 1.3.14.22). Indeed, as Wade Robison observes, Hume seems to *identify* the impression of efficacy with the “mind’s *determination* to conceive of or believe in the occurrence of *b* upon the perception of *a*.”⁹

Hume doesn’t just declare that associations of ideas are paradigms of efficacy. He goes so far as to conclude that efficacy isn’t in bodies: “this customary transition is, therefore, the same with the power and necessity; which are consequently qualities of perceptions, not of objects” (*THN* 1.3.14.24). People can use words as they please, but “when we transfer the determination of the thought to external objects, and suppose any real intelligible connexion betwixt them” we “make the terms of power and efficacy signify something, of which we have a clear idea” at the same time that we make them signify something “which is incompatible with those objects, to which we apply it” (*THN* 1.3.14.27).¹⁰

The belief to the contrary is merely the result of projection. “‘Tis a common observation,” Hume writes, “that the mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses” (*THN* 1.3.14.25). According to Peter Kail’s treatment of this passage and related ones, “Hume’s strategy is best read as allowing the possibility of necessary connections [including ones between external objects--MJ], and its main concern is to show that we are not receptive to any such features.”¹¹ Kail’s paper is good, but it’s better as charitable reconstruction than as a report of what Hume wrote. Though Kail has many illuminating

⁹ Wade L. Robison, ‘Hume’s Causal Scepticism,’ in *David Hume: Bicentenary Papers*, edited by G.P. Morice (Edinburgh: Edinburgh University Press, 1977), 156-66, at 158.

¹⁰ P. Kyle Stanford, ‘The Manifest Connection: Causation, Meaning, and David Hume’ [‘Connection’], *Journal of the History of Philosophy* 40 (2002): 339-60, at 346-47

¹¹ P.J.E.Kail, ‘Projection and Necessity in Hume’ [‘Necessity’], *European Journal of Philosophy* 9 (2001): 24-54, at 34.

examples of projection in his paper, he doesn't directly cite Hume's own illustration in the *Treatise*: "Thus as certain sounds and smells are always found to attend certain visible objects, we naturally imagine a conjunction, even in place, betwixt the objects and qualities, tho' the qualities be of such a nature as to admit of no such conjunction, and really exist no where" (*THN* 1.3.14.25). That is, efficacy is like smell and sound which we imagine to be conjoined, spatially and otherwise, to external objects, even though they can't be.¹² As Kyle Stanford argues,¹³ if the analogy isn't inept, its point must be that when we attribute any of efficacy, sound, or smell to external objects, we attribute qualities to them that they cannot possibly bear.

If external objects can't have efficacy in them, there's no point in looking for it there. In our inquiry after the first principles behind ordinary causes, Hume tells us,

We wou'd not willingly stop before we are acquainted with that energy in the cause, by which it operates on its effect; that tie, which connects them together; and that efficacious quality, on which the tie depends. This is our aim in all our studies and reflections: and how must we be disappointed, when we learn, that this connexion, tie, or energy lies merely in ourselves, and is nothing but that determination of the mind, which is acquir'd by custom, and causes us to make a transition from an object to its usual attendant, and from the impression of one to the lively idea of the other? Such a discovery not only cuts off all hope of ever attaining satisfaction, but even prevents our very wishes; since it appears, that when we say we desire to know the ultimate and operating principle, as something, which resides in the external object, we either contradict ourselves, or talk without a meaning (*THN* 1.4.7.5)

Not only is there no hope of the finding efficacy in external objects, but even asking after the efficacy of external causes is meaningless or worse.

So we may map Hume's views onto Malebranche's. The vulgar believe that ordinary objects have power in them, but the philosopher sees that the locus of efficacy is elsewhere.

¹² Hume drops a footnote (*THN* 1.14.25n32) to *Treatise* 1.4.5, where he argues that smell, sound, and taste don't exist in external objects.

¹³ Stanford, 'Connection,' 353-55.

According to Malebranche, true efficacy lies in God; according to Hume, in the mind. On Malebranche's account, we falsely attribute efficacy to bodies because we are deceived by our senses. On Hume's account, we falsely attribute efficacy to bodies because "the mind has a great propensity to spread itself on external objects" (*THN* 1.3.14.25)—a phrase that he seems to have lifted from Malebranche.¹⁴

One tendency in recent Hume scholarship has been to soft-pedal such subjectivist elements. I don't myself think that they are an important part of Hume's considered view of the late 1740s. Still, for the purposes of the developmental story that I want to tell, it's important to see where these elements came from and to see where they went. If the story is plausible, if it explains otherwise inexplicable texts, and if it solves otherwise intractable problems, then my *prima facie* case that Hume is offering a serious piece of subjectivist metaphysics in *Treatise* 1.3.14 becomes decisive.

2.2 Hume's Second Definition of Causation

We've been examining what Hume calls 'efficacy,' a quality that he says is restricted to the soul and can't exist in external objects. His definition of cause, in contrast, is intended to apply to what ordinary people call causes, including external objects. In this section, I want to discuss Hume's derivation of his second definition of cause and ask whether he intends the definition to apply to associations of ideas. My larger goal is to show that his problematic answers to these questions are repeated in answer to analogous questions in his chapter on personal identity and that these produce the difficulties he bemoans in the Appendix.

¹⁴ Kail, 'Necessity,' 52n87.

In Book 1, Part 3 of the *Treatise*, Hume sets himself the project of explaining causation “fully” (*THN* 1.3.2.3). In this pursuit, he initially argues that there are three components of the relation of cause and effect: contiguity, succession, and necessary connection (*THN* 1.3.2.6-11).¹⁵ Since he has trouble finding the impression of necessary connection, he concludes that we don’t have a good understanding of that relation. The problem is difficult enough that he declares,

‘Tis necessary for us leave the direct survey of this question concerning the nature of that *necessary connexion*, which enters into our idea of cause and effect; and endeavour to find some other questions, the examination of which will perhaps afford a hint, that may serve to clear up the present difficulty (*THN* 1.3.2.13).

Among those questions, Hume asks, “why we conclude, that such particular causes must *necessarily* have such particular effects; and what is the nature of that *inference* we draw from the one to the other, and of the *belief* we repose in it?” (*THN* 1.3.2.15). Hume’s questions hang together through his assumption that necessary connection, whatsoever it may be, is what makes us infer the existence of an effect from the appearance of a cause and *vice versa*. A cause is an object contiguous and precedent to another, with a blank for whatever makes us believe in the existence of the effect when we are presented with the cause.

After his investigation of causal inference, Hume concludes that its distal cause of our belief is the constant conjunction of objects of one sort with objects of another sort. Its proximal cause is the association that custom carves between the impression of one and the lively idea of the other. He fills in the blanks accordingly.¹⁶ When he fills in the blank with constant conjunction, the distal cause of our belief, his first definition of cause runs, “An object precedent and contiguous to another, and where all the objects resembling the former

¹⁵ In a footnote (*THN* 1.3.2.6), he points to *Treatise* 1.4.5 and his final judgment that contiguity isn’t essential to causation.

¹⁶ McRae, ‘Meaning,’ 490-91.

are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter" (*THN* 1.3.14.31). When he fills in the blank with the psychological association, the proximal cause of our belief, his second definition runs: "A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other" (*ibid.*).

This account of the derivations of Hume's definitions explains his remark that the two definitions "are only different, by their presenting a different view of the same object" (*ibid.*). The first takes a long view of the source of the inference, and the second takes an introspective view. This difference of view leads to a genuine difference in definition.

In understanding the relation between these definitions, I find it helpful to look back at his reply to the shocked interlocutor. Not only does the response give a full and forthright answer to the objection that Hume's account makes causation subjective, but it also illuminates the two definitions of causation.

In reply to the objection that "the operations of nature are independent of our thought and reasoning," Hume begins by conceding that contiguity, succession, and the repetition of those relations is independent of any observer,

I allow it; and accordingly have observ'd, that objects bear to each other the relations of contiguity and succession; and that like objects may be observ'd in several instances to have like relations; and that all this is independent of, and antecedent to the operations of the understanding. (*THN* 1.3.14.28)

Since these are the elements of Hume's first definition of causation, it follows that causation in that sense obtains independently of any observer. Hume goes on to reject the completeness of such an account: "if we go any farther, and ascribe a power or necessary connexion to these objects; this is what we can never observe in them, but must draw the idea of it from what we feel internally in contemplating them" (*THN* 1.3.14.28). If we want

a complete account of causation, one that includes necessity, we need to appeal to associations between our perceptions.¹⁷ Those are the exemplars of efficacy from which we draw the corresponding idea. Hume implies that his first definition of causation is incomplete in a way that his second definition is not.

In *Treatise* 2.3.2, Hume seems to treat the definitions as being on a par: “I define necessity two ways, conformable to the two definitions of *cause*, of which it makes an essential part. I place it either in the constant union and conjunction of like objects, or in the inference of the mind from the one to the other” (§4). Flatly read, this contradicts his earlier claim that ascriptions of necessity require an appeal to inner associations. It *is* possible to ascribe necessity to bodies without drawing on our inner feeling of customary inference. Hume has done it.

We can still capture the spirit behind his assertion that ascriptions of necessity as a genuine relation between objects require reference to an inner impression of association by attending to his claim that the first definition characterizes causation as a philosophical relation and the second as a natural relation (*THN* 1.3.14.31). Hume introduces the distinction between the two kinds of relation by pointing to differences in the way that philosophers and ordinary people use the word ‘relation’.¹⁸ In philosophy, any two things may be related along any dimension. For example, I stand in the *larger than* relation to a pebble on the shores of the Baltic. In ordinary language, two things must have more to do with one another count as related:

¹⁷ See Norman Kemp Smith, *The Philosophy of David Hume [Philosophy]*, with a new introduction by Don Garrett (Basingstoke, Hampshire: Palgrave Macmillan, 2005), 91-93 and Don Garrett, *Cognition and Commitment in Hume’s Philosophy [Cognition]* (Oxford: Oxford University Press, 1997), 99-100.

¹⁸ Eric Schliesser, ‘Two Definitions of ‘Cause,’ Newton, and the Significance of the Humean Distinction between Natural and Philosophical Relations,’ *The Journal of Scottish Philosophy* 5 (2007): 83-101, at 92.

distance will be allowed by philosophers to be a true relation, because we acquire an idea of it by the comparing of objects: But in a common way we say, *that nothing can be more distant than such or such things from each other, nothing can have less relation*; as if distance and relation were incompatible (THN 1.1.5.1).

In this ordinary way of speaking, I'm related to images of me, to my clothing, and to my sister, but not to images of you, to your clothing, or to your sister. Hume's analysis of the ordinary notion of relation is that two things are related in this sense, if the idea of one is associated with the idea of the other (THN 1.3.6.16, 1.3.14.31).¹⁹

We may, Hume grants, say that two objects are causally related when they stand in objective and mind-independent relations of contiguity, succession, and resemblance to other pairs of contiguous, successive objects. In order to 'go any further' in these ascriptions, we need to draw on our inner associations of ideas between the two. His point, I'm pretty sure, is that when two objects satisfy the first definition of causation, they are related by a relation of cause and effect, but only in the sense that I'm related to a pebble near the Baltic. A genuine, natural relation requires an association between the idea of the cause and the idea of the effect.²⁰

We may thus distinguish three Humean grades of relatedness: *philosophically related*, if someone might compare two things along some dimension, *naturally related*, if the thought of one leads along a certain path to the thought of another, and *really connected*, if the impression of the one would allow us to infer the existence of the other. Sometimes Hume appeals to his definition of causation as a philosophical relation (e.g. THN 1.3.15.1, 1.3.15.5, 1.4.5.32).²¹

¹⁹ Thomas J. Richards, 'Hume's Two Definitions of 'Cause', *The Philosophical Quarterly* 15 (1965): 247-53, at 248-49.

²⁰ Two objects that meet the first definition of cause and effect will always be naturally related, since they will be contiguous, but Hume is more interested in whether they are connected by the natural relation of cause and effect.

²¹ I owe the references to a referee.

Sometimes he doesn't. Hume gets into his regress by trying to analyze too many instances of causation as natural relations.

Though Hume admits a certain degree of mind-dependence in his account of causation, the admission is less dramatic than the earlier parallels with Malebranche might have suggested. We may gild and stain the world with powers projected from our associations of ideas, but the spatio-temporal arrangement of objects doesn't depend on anyone's mind. Though the natural relation between A and B may depend on the association of ideas in some observer, the fact that A touches and precedes B does not. The trouble lies elsewhere.

2.3 Does the Second Definition Apply to Associations of Ideas?

As the second definition stands in the *Treatise*, two associations of perceptions must obtain for the objects they represent to be causally related: "that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other." The first kind of association connects ideas and regulates us in thinking things over, as when the thought of Henry VII leads us to think of Henry VIII (*THN* 1.1.4.3) or when the thought of Sarpedon's wounds leads us to think of the pain that they caused (*EHU* 3.3). This, along with principles governing ideas of resembling objects and ideas of contiguous objects, is one of three principles of association between ideas that Hume describes in *Treatise* 1.1.4 and *Enquiry* §3. I'll sometimes call such associations between ideas 'contemplative associations.'

The second kind of association connects the impression that we get from viewing a possible cause with the anticipatory belief in what the cause would or will bring about, as when looking at a river causes the lively idea of drowning (*THN* 1.3.8.13) or when looking at

one billiard ball moving towards another makes us anticipate that the second ball will soon be moving (*EHU* 5.11). This kind of association of impression and lively idea also works in the other direction, so that from the impression of an effect, we form a lively idea of its cause; for example, from the impressions of ink on paper we form a lively idea that Caesar was killed on March 15th (*THN* 1.3.4.2). I'll call associations between impressions and lively ideas 'lively associations.'

On the face of it, associations between perceptions are themselves causal relations. Does Hume believe that his second definition of causation offers a proper analysis of lively and contemplative associations? For example, does he believe that the idea of Henry VII is associated with the idea of Henry VIII if and only if the impression of the idea of Henry VII is associated with the lively idea of the idea of Henry VIII and the idea of the idea of Henry VII is associated with the idea of the idea of Henry VIII? Such an application would be unwieldy at best.

According to Aryeh Botwinick, the second definition is not supposed to apply to the associations in the *definiens*. Rather, "The 'felt necessity' stands in a different relation to us than other causal connections."²² At first glance, it seems as if Botwinick must be right. Otherwise, behind every association of perceptions must stand another association of distinct perceptions. Not only is this contrary to the psychological phenomena, since we don't find a higher-level association of ideas lurking behind every association of ideas, but it would give rise to a vicious infinite regress.

Moreover, both Hume's vigorously stated thesis that the 'energy' of causes is restricted to the mind and his projective diagnosis of his opponents' bias would be quite

²² Aryeh Botwinick, *Ethics, Politics, and Epistemology: A Study in the Unity of Hume's Thought* (Washington D.C.: University Press of America, 1980), 98, see also Antony Flew, *Hume's Philosophy of Belief: A Study of his First Inquiry [Belief]* (New York: The Humanities Press, 1961), 122-23.

puzzling if Botwinick is wrong and Hume doesn't believe that causal connections in the mind have a special metaphysical status. The efficacious quality binding causes and effect, he tells us, "lies merely in ourselves, and is nothing but that determination of the mind, which is acquir'd by custom, and causes us to make a transition from an object to its usual; attendant, and from the impression of one to the lively idea of the other" (*THN* 1.4.7.5). These doctrines imply, consonantly with Botwinick's interpretation, that the kind of efficacy that connects perceptions together is foundational and underlies the connections between all the other, second-rate causes.

On the other hand, Hume explicitly says that his account of causation applies to perceptions in the mind. He declares himself "ready to convert my present reasoning into an instance of it, by a subtility, which it will not be difficult to comprehend" (*THN* 1.3.14.28). In ordinary cases, he writes, an object "conveys to the mind a lively idea of that object, which is usually found to attend it; and this determination of the mind forms the necessary connexion of those objects" (*THN* 1.3.14.29). This is what I've called a lively association. In application to this lively association, Hume believes that his account triumphs again: "in that case the impression is to be considered as the cause, and the lively idea as the effect; and their necessary connexion is that new determination, which we feel to pass from the idea of the one to that of the other" (*ibid.*). On this account, the lively association rests on a contemplative association.

Hume's reason for passing the buck from the lively association to the contemplative one is that "the uniting principle among our internal perceptions is as unintelligible as that among external objects, and is not known to us any way other than by experience" (*ibid.*). So it seems that from the fact that we don't perceive a real intelligible connection between the impression and the lively idea, he is willing to conclude that efficacy isn't a quality that

belongs to those perceptions, but rather to the ideas of those perceptions. This inference only makes sense if Hume assumes that efficacy requires perceptible, intelligible connections.

Parity of reasoning would seem to commit him applying the same analysis to the contemplative association. Early in the *Treatise*, he describes associations of ideas as

a kind of ATTRACTION, which in the mental world will be found to have as extra-ordinary effects as in the natural, and to shew itself in as many and as various forms. Its effects are every where conspicuous; but as to causes, they are mostly unknown, and must be resolv'd into *original* qualities of human nature, which I pretend not to explain (*THN* 1.1.4.6).

So, he doesn't think of contemplative associations as examples of self-explanatory intelligible connections either.

Hume implies that behind a lively association lies a contemplative one. If he thinks that behind every contemplative association lies another contemplative association, then he seems to be stuck with an infinite regress. Hume makes three assumptions: one, when we feel a connection between two objects without perceiving it, the feeling arises from the association of the corresponding ideas; two, perceiving a connection between two objects requires perceiving an intelligible connection; and, three, we never perceive intelligible connections between any two objects. These assumptions threaten to generate a wild-goose chase, where we are always on the lookout for intelligible connections just over the horizon in the next association of ideas.

2.4 Garrett's Way Out

In his justly celebrated treatment of Hume's definitions of cause, Don Garrett worries that circularity threatens the second definition.²³ This isn't exactly my concern, since it seems to me that there can be good definitions where the *definiendum* occurs in the *definiens*, as we find

²³ Garrett, *Cognition*, 100.

in inductive definitions. Rather, my worry is that Hume's definition, if anything falls under it, entails the existence of an infinite stack of associations of ideas, each resting on another.

Even so, Garrett's treatment of Hume's second definition offers a nice way out of my worry. Garrett distinguishes subjective and idealized readings of Hume's two definitions of cause. According to the subjective reading, the second definition of cause

provides a 'subjective,' person-relative sense of 'cause', according to which one object is a cause of another object for a particular person if and only if it is prior (and, perhaps, contiguous) to the other object and psychologically associated with it, in the way that C2 specifies, *for that person*.²⁴

On this interpretation of the second definition, a causal connection between two objects requires that some actual mind associates the corresponding perceptions.

According to the idealized (or 'absolute') reading, the observing mind in the second definition of causation is "an idealized mind or spectator—for example, one who accurately views all and only representative samples, has a well-developed human inferential mechanism, and suffers from no interfering biases such as those deriving from religion or eccentricities of the imagination."²⁵ On this interpretation, two objects are related as cause and effect if they touch, one precedes the other, and a properly situated idealized mind would associate the corresponding perceptions. As Garrett argues,²⁶ if hypothetical observers suffice to connect cause and effect, then the second definition would be equivalent to the first. I would go further and say that on this reading, the second definition *reduces* to the first. By tacit stipulation, the idealized hypothetical observer viewing constant conjunctions of objects associates the idea of the one with the idea of the other. The psychological patterns in hypothetical observers would be decorations, not load-bearing elements.

²⁴ *ibid.*, 108.

²⁵ *ibid.*, 108-09.

²⁶ *ibid.*

If the inferences of a hypothetical observer suffice to bind events together with necessity, then there is no longer any phenomenological problem or a problem about infinite regresses with the principles of association. To say that the idea of A leads to the idea of B is just to say that ideas like A are constantly conjoined with ideas like B, and that, if a hypothetical observer were to view that constant conjunction, then he would associate the idea of the idea of A with the idea of the idea of B. It's part of the hypothesis that an association will form in the observer in these circumstances, and there's no need to apply this analysis to the contemplative associations in the hypothetical observer's mind. The hypothetical observer adds nothing to the pattern of constant conjunction.

Garrett observes that if we consider Hume's second definition of cause as "referring to an idealized mind," then it is compatible with the existence of hidden springs and principles and with Hume's normative rules for judging cause and effects.²⁷ In addition, on the idealized reading, the second definition

implies neither (i) that objects observed to be conjoined in unrepresentative samples are always real causes, nor (ii) that the existence or nonexistence of a causal relation is relative to individual minds, nor (iii) that there would be no causation at all unless there were minds.

For these reasons among others, Garrett declares, "I have little doubt that [Hume] would ultimately prefer the absolute reading, at least for most purposes."²⁸

Garrett's caveat, 'for most purposes,' ought to be heeded. On the idealized reading of the second definition, the subjectivist elements in Hume's discussion of causation drop out. On the idealized reading, Hume hasn't seriously engaged in Malebranche's project of finding the true engine of apparent causes, and he hasn't restricted efficacy to the soul in any metaphysically interesting sense. On this reading, causal connections don't depend on any

²⁷ *ibid.*, 112.

²⁸ *ibid.*, 109.

actual associations of perceptions in any actual mind. Constant conjunction does all the work.

If my goal were to provide a charitable reconstruction of Hume's account of causation, the trivialization of the subjective elements in his treatment wouldn't bother me. A metaphysics that makes every causal connection in the world depend on some actual association of ideas seems to me to face insuperable difficulties. But my goal is *not* to provide a clean reconstruction of Hume's theory, but rather to find the sources of the contradiction that he laments in the Appendix. For that end, I want to heighten the contradictions in Hume's analysis of causation, not to paper them over.

In the next part of my paper, I'll discuss the explicit connection that Hume draws between his analysis of causation and his analysis of personal identity. In the fourth part, I'll show how the general problem of the infinite regress of actual associations of ideas arises in the special case of personal identity. In the final part, I'll explain how Hume resolved the difficulty in the *Enquiry* by moving entirely and consistently to Garrett's absolute reading of the second definition. This cuts off the infinite regress and pushes Hume towards a less mind-dependent account of causation. Eventually, Hume takes Garrett's way out, but not before his application of the second definition of causation to personal identity gets him tangled up and bewildered.

3 Associations of Ideas in Hume's Theory of Personal Identity

3.1 People as Bundles of Perceptions

Having begun *Treatise* 1.4.6 ("Of personal identity") with arguments that our perceptions don't inhere in a simple enduring substance, Hume ends the section with an inquiry into the relation that binds a person's perceptions together. In this inquiry, he imports wholesale the

methods and assumptions which he applied to his analysis of causation. No real connections are perceptible between perceptions, so personal identity must be analyzed as resting on the associations of ideas which represent those perceptions. Hume ties his accounts of causation and personal identity together so tightly that he finishes by concluding that causation constitutes personal identity.

One of Hume's arguments against the traditional doctrine that a person's perceptions inhere in an enduring substance is that we have no impression of an enduring, simple substance. He reports, "For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure" (*THN* 1.4.6.3). He concedes,

If any one upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself*, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continu'd, which he calls *himself*; tho' I am certain there is no such principle in me. But setting aside some metaphysicians of this kind, I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement (*ibid.*).

Since we perceive nothing inside of ourselves besides perceptions, persons are bundles of perceptions.

In the Abstract, Hume offers another version of this argument, with a nominalistic introduction. Considering the later argument will help us reconstruct the assumptions behind the earlier one. He writes,

Des Cartes maintained that thought was the essence of mind; not this thought or that thought, but thought in general. This seems to be absolutely unintelligible, since every thing that exists, is particular: and therefore it must be our several particular perceptions, that compose the mind. I say, *compose* the mind, not *belong* to it. The mind is not a substance in which the perceptions inhere. That notion is as unintelligible as the *Cartesian*, that thought or perception in general is the essence of the mind. We have no idea of substance of any kind, since we have no idea but what is derived from

some impression, and we have no impression of any substance either material or spiritual (*THN Abstract* ¶28).

Either a person's perceptions compose the mind or the mind is a substance in which they inhere. If the mind is a substance in which they inhere, then we have some idea of substance. But we have no impression of substance, so we have no idea of substance.²⁹ So a person's perceptions compose a mind. (These perceptions are particular and not general, since everything is particular.)

If a person's perceptions compose a mind, then there's some relation that ties them together. According to Hume, figuring out the nature of this relation is one of the hardest problems in metaphysics:

'Tis certain there is no question in philosophy more abstruse than that concerning identity, and the nature of the uniting principle, which constitute a person. So far from being able by our senses merely to determine this question, we must have recourse to the most profound metaphysics to give a satisfactory answer to it (*THN* 1.4.2.6).

So, to sum up the assumptions behind Hume's inference, since we don't have an impression of substance, we don't have an idea of substance, and since we don't have an idea of substance, our perceptions don't inhere in a substance, and since our perceptions don't inhere in a mind, they compose a mind, and since they compose a mind they are bound together by some relation. Figuring out the nature of this relation requires deep thought.

Notwithstanding such texts, many commentators have agreed with Sybil Wolfram, who writes, "where Hume is supposed to have been seeking to solve the problem of personal identity he was trying to discover how the mistaken conviction that persons have

²⁹ John Bricke. *Hume's Philosophy of Mind [Hume's Mind]* (Princeton: Princeton University Press, 1980), 61-63 has useful discussion of this argument.

identity can have arisen.”³⁰ This view is common enough among commentators³¹ that I should explain why I reject it.

The only reason to deny that Hume is offering a positive account of personal identity would be if any such account would contradict his doctrine that identity in the strict and philosophical sense requires changelessness. On close examination, we see that it doesn't. Hume has two senses of 'identity,' only one of which requires qualitative identity over time. He distinguishes the two senses at 1.1.5, writing, under the heading of *identity*, “this relation I here consider as apply'd in its strictest sense to constant and unchangeable objects; without examining the nature and foundation of personal identity, which shall find its place afterwards” (*THN* 1.1.5.4). That place is *Treatise* 1.4.6. For Hume, personal identity is not a species of identity in the strict and philosophical sense.

Hume, as Norman Kemp Smith writes, “supplements his rigorous view of identity as exhibited in simples, and as therefore being always absolute, with a less strict type of identity, proper to certain complex entities. The identity of the self, he is arguing, is of the latter type.”³² Lawrence Ashley and Michael Stack and Susan Mendus observe that Hume believes that a planet may continue to possess “imperfect identity” after the addition of a mountain (*THN* 1.4.6.9).³³ They show that for Hume personal identity is a species of imperfect

³⁰ Sybil Wolfram, 'Hume on Personal Identity,' *Mind* 83 (1974): 586-93, at 589.

³¹ *inter alia*, Terence Penelhum, *Themes in Hume: The Self, the Will, Religion* [Themes] (Oxford: Clarendon Press, 2000), 28-32, Norman Melchert, 'Hume's Appendix on Personal Identity' [Appendix], *Philosophy Research Archives* 1 (1975): 323-35, at 328-29, Kenneth P. Winkler, “All is Revolution in Us: Personal Identity in Shaftesbury and Hume” [Revolution] *Hume Studies* 26 (2000): 1-40, at 18-19, and Donald Ainslie, 'Hume's Reflections on the Identity and Simplicity of Mind' [Reflections], *Philosophy and Phenomenological Research* 62 (2001): 557-78, at 563.

³² Kemp Smith, *Philosophy*, 97.

³³ Lawrence Ashley and Michael Stack, 'Hume's Theory of the Self and its Identity' [Self], *Dialogue* 13 (1974): 239-54, at 244-45 and Susan Mendus, “Personal Identity: The Two Analogies in Hume” [Analogies], *The Philosophical Quarterly* 30 (1980): 61-68, at 66-67. See also James Noxon, “Senses of Identity in Hume's *Treatise*,” *Dialogue* 8 (1969): 367-73, at 370-73 and Ainslie, 'Reflections,' 563.

identity.³⁴ He would hardly write, “as the same individual republic may not only change its members, but also its laws and constitutions; in like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity,” (*THN* 1.4.6.19) if he had in mind a notion of identity that requires changelessness.

According to Terence Penelhum, Hume’s belief in a widespread error about strict identity is incompatible with a picture of the mind as an enduring construction out of perceptions, since “logical” constructions cannot be founded on error:

We think the mind has strict identity, when in reality it does not have identity at all; so the mind of common sense belief cannot be a logical construction out of its perceptions, since if it were, such a false judgement could not emerge as the result of translation from the language of perception to that of common sense.³⁵

‘Natural construction’ may be a better term here. In the Appendix, Hume summarizes his positive project in *Treatise* 1.4.6 before he expresses his unhappiness with his account. He describes the relevant relation between all our perceptions as “the principle of connexion, which binds them together, and makes us attribute to them a real simplicity and identity” (*THN Appendix* ¶20). Notice that Hume attributes two effects to the principle of connection. First, it binds our perceptions together. Second, it misleads us into false judgments of simplicity and strict identity. Hume, unlike Penelhum, believes that associations of ideas can do both.³⁶

It may be objected that Hume himself tells us that we make an ontological error when we make mistaken judgments about identity:

³⁴ Ashley and Stack, ‘Self,’ 245-47, Mendus, ‘Analogies,’ 68.

³⁵ Penelhum, *Themes*, 113.

³⁶ Penelhum also remarks, “I still incline to take him literally when he tells us that ‘(t)he identity, which we ascribe to the mind of man, is only a fictitious one’ (*TI*.iv.6 259), and not him as saying merely that this identity is *factitious*” (*ibid.*, 114). Hume does believe that imperfect identity is second-rate next to strict identity, but we should keep in mind ‘fictitious’ once bore the meaning of artificial (*OED s.v.* ‘fictitious,’ Def. 1a).

when we attribute identity, in an improper sense, to variable or interrupted objects, our mistake is not confin'd to the expression, but is commonly attended with a fiction, either of something invariable and uninterrupted, or of something mysterious and inexplicable, or at least with a propensity to such fictions (*THN* 1.4.6.7).

The mistake that he has in mind, however, is not that of believing that changing things might endure, but rather that of believing that there is something to enduring, changing things beyond “a succession of parts, connected together by resemblance, contiguity, or causation” (*ibid.*). We err not by judging that trees can endure the shedding of their leaves, but rather by imagining “something unknown and mysterious, connecting their parts, besides their relation; and this I take to be the case with regard to the identity we ascribe to plants and vegetables” (*THN* 1.4.6.6).³⁷ Hume drops a disbelieving reference (*THN* 1.4.6.6n50) to Shaftesbury, who thought that considerations of identity require us to posit that trees have an enduring “peculiar nature” in addition to their wood, bark, and leaves (*Characteristics* 300).³⁸

Hume believes that changing things may endure through relations that obtain between diverse parts. Indeed, in Part 4 of the *Dialogues Concerning Natural Religion*, he implies that change is essential to minds: “A mind whose acts and sentiments and ideas are not distinct and successive, one that is wholly simple and totally immutable, is a mind which has no thought, no reason, no will, no sentiment, no love, no hatred; or, in a word, is no mind at all” (*DNH* 61). The shock is that he believes that minds and other entities can endure without identity, strictly so-called. The shock would be greater if his strict notion of identity weren't so straitened and extraordinary. Mendus convincingly argues that we ought to respect Hume's comparisons between persons and living things (*THN* 1.4.6.15) and between

³⁷ See Ashley and Stack, 'Self,' 245-46. 252-53.

³⁸ See Winkler, 'Revolution,' 13-15.

persons and commonwealths (*THN* 1.4.6.19).³⁹ Hume presents these changing and enduring objects as models for the metaphysics of personal identity.

3.2 Hume's Second Argument for His Theory of Personal Identity

Let me continue on the assumption that Hume is interested in the metaphysics of personal identity and not just in the source of a false belief. He offers two arguments for his theory. The first is that the account “which has so successfully explain'd the identity of plants, and animals, and ships, and houses” ought to work as well for “the mind of man” (*THN* 1.4.6.15). He then offers a second argument, which he considers more rigorous and more direct (“closer and more immediate” *ibid.*). In this argument, he links causation and personal identity and argues that the same considerations which show that causation “resolves itself into a customary association of ideas” (*THN* 1.4.6.16) also show that personal identity should be analyzed in the same way.

Hume begins his second argument as follows:

But lest this argument shou'd not convince the reader; tho' in my opinion perfectly decisive; let him weigh the following reasoning, which is still closer and more immediate. 'Tis evident, that the identity, which we attribute to the human mind, however perfect we may imagine it to be, is not able to run the several different perceptions into one, and make them lose their characters of distinction and difference, which are essential to them. 'Tis still true, that every distinct perception, which enters into the composition of the mind, is a distinct existence, and is different, and distinguishable, and separable from every other perception, either contemporary or successive. But, as, notwithstanding this distinction and separability, we suppose the whole train of perceptions to be united by identity, a question naturally arises concerning this relation of identity; whether it be something that really binds our several perceptions together, or only associates their ideas in the imagination. That is, in other words, whether in pronouncing concerning the identity of a person, we observe some real bond among his perceptions, or only feel one among the ideas we form of them? (*ibid.*)

³⁹ Mendus, 'Analogies,' 63-68.

The argument concerns the relation that we suppose unites “the whole train of perceptions.” Hume begins with the obvious facts that this relation can’t possibly make “the several different perceptions into one” nor could it “make them lose their characters of distinction and difference.” These facts imply that the relation is not identity in Hume’s strict and philosophical sense.

“Notwithstanding this distinction and separability,” he continues, “a question naturally arises concerning this relation of identity; whether it be something that really binds our several perceptions together, or only associates their ideas in the imagination.” What does this question mean? We have already seen that for Hume real connections bind their relata in an intelligible and context-independent manner. What is it for a relation to associate ideas of perceptions in the imagination? We need to answer this question before we can understand the purpose of Hume’s second argument, and we need to understand its purpose before we can understand its structure.

The solution may be found in Hume’s reference, a little later in the paragraph, to having “prov’d at large” that “the union of cause and effect, when strictly examin’d, resolves itself into a customary association of ideas.” The lengthy proof to which he refers must surely be his discussion of causation in Book 1, Part 3, and the relevant resolution of cause and effect into a custom-based association of ideas must be his second definition of cause and effect.

Let me call relations between A and B which have the form, *the idea of A is psychologically associated with the idea of B*, ‘merely attributed relations.’ When I say that a relation has this form, I don’t just mean that it provokes an association when observed, but that the association of ideas is essential to the definition or analysis of the relation. Thus, resemblance and contiguity provoke associations of ideas, but they are mind-independent

relations on Hume's view, and associations of ideas play no role in their analysis. By the 'idea of A', I mean the idea that represents A. The relata A and B may themselves be impressions, distinct ideas, or anything else.

For whatever reason, Hume's reference to his analysis of cause and effect "into a customary association of ideas" omits any reference to lively associations. Setting lively associations aside, Hume's second definition of causation analyzes it as a merely attributed relation. When he asks whether identity is a relation that associates ideas of perceptions in the imagination, he is asking whether personal identity is a merely attributed relation. Thus, on my interpretation, the question that he says naturally arises is whether our perceptions hang together in a context-independent way or whether they are connected indirectly through associations of the corresponding ideas.

After asking whether the relation of identity is "something that really binds our several perceptions together, or only associates their ideas in the imagination," Hume restates the question as follows: "That is, in other words, whether in pronouncing concerning the identity of a person, we observe some real bond among his perceptions, or only feel one among the ideas we form of them?" The original question asks whether personal identity is a real bond between perceptions. The second asks whether we observe a real bond between our perceptions. Hume's characterization of the second as a paraphrase of the first shows that he assumes that if a relation really bound our perceptions together, then we would be able to observe it.

Hume offers us the following argument in answer to the problem of whether personal identity is a real connection between perceptions:

This question we might easily decide, if we wou'd recollect what has been already prov'd at large, that the understanding never observes any real connexion among objects, and that even the union of cause and effect, when strictly examin'd, resolves itself into a customary association of ideas. For

from thence it evidently follows, that identity is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect upon them (ibid.).

To begin at the end of the argument, the conclusion is that the relevant identity “is merely a quality, which we attribute to [different perceptions], because of the union of their ideas in the imagination, when we reflect upon them.” Which is to say, in my terminology, that Hume’s conclusion is that personal identity is a merely attributed relation.

The argument for this conclusion runs as follows:

1. We never observe any real connection between objects (Premise).
2. If identity is a real connection between perceptions, then we sometimes observe it (premise).
3. So, identity is not a real connection between perceptions (1, 2).
4. Identity holds between perceptions and is either a real connection or a merely attributed relation arising from the association of ideas (Premise).
5. So, identity is a merely attributed relation arising from the association of ideas (3,4).

The first premise is explicit, both here and as the second problematic principle in the Appendix (*THN Appendix ¶21*). That he assumes the second premise is clear from his initial paraphrase of his problem. Moreover, Hume concludes that personal identity is not a real connection, and Premise 2 is the straightforward way of making the imperceptibility of real connections relevant to that conclusion. Without it or something like it, it would just be a non sequitur to say that the unreality of the relevant relation of identity follows from the premise that we never observe a real connection between distinct objects.

From these premises, Hume draws the sub-conclusion that identity is not a real connection between perceptions.⁴⁰ Notice that his argument that identity is not a real connection between perceptions wouldn’t make any sense if he had the strict and philosophical sense of identity in mind. Appealing to the general and controversial claim

⁴⁰ Loeb, ‘Relations,’ 219.

that we never perceive any real connections between distinct objects would be pointlessly weak; he could just appeal to the obvious fact that distinct perceptions are not qualitatively similar.

Proposition 4 gets us to the desired conclusion that personal identity is a merely attributed relation. The dichotomy is explicit in Hume's initial presentation of the problem ("a question naturally arises concerning this relation of identity; whether it be something that really binds our several perceptions together, or only associates their ideas in the imagination"). He justifies this premise only by referring back to his discussion of causation, where he had also managed to move from the imperceptibility of a relation to analyzing it as resting upon associations of perceptions. No real connections are perceptible between perceptions, so personal identity must be analyzed as resting on contemplative associations of ideas of those perceptions.

It is jarring to describe identity as a relation that holds between distinct perceptions. Let me make two remarks in defense of Hume's usage. First, he calls the relevant relation "imperfect identity" (*THN* 1.4.6.9) and says that when we apply the word 'identity' to changing objects, we do so "in an improper sense" (*THN* 1.4.6.7), so a certain amount of loose talk is to be expected. Second, when Hume wrote, the problem of personal identity had "become so great a question in philosophy, especially of late years in *England*" (*THN* 1.4.6.15). Locke, the stimulus of this debate, treated principles of identity as principles that describe how an enduring thing can be composed of different constituents over time. For example, according to Locke, "wherein the Identity of the same *Man* consists" is "nothing but a participation of the same continued Life, by constantly fleeting Particles, in succession vitally united to the same organized Body" (E 2.27.6). This principle of identity doesn't

answer the question ‘why is a man self-identical?’⁴¹ Rather Locke’s account, if it succeeds, tells us under what conditions various particles come to compose an enduring man. With certain supplemental assumptions, answers to questions about when and where a certain man is and whether he is the same man as so-and-so will come in train. In his approach to imperfect identity, Hume is imitating Locke.⁴²

3.3 The Consequences of the Argument

One virtue of my interpretation of this argument is that it fits nicely with the text that follows it. Hume begins the next paragraph by writing, “The only question, therefore, which remains, is, by what relations this uninterrupted progress of our thought is produc’d, when we consider the successive existence of a mind or thinking person” (*THN* 1.4.6.17). He then considers his three principles of association between ideas as enumerated in *Treatise* 1.1.4 and asks what role they might play in constituting personal identity. This is just the detail that we would expect Hume to specify if his conclusion is that personal identity is a merely attributed relation that arises out of an association of ideas.

Of his three principles of association, Hume decides that causation does most of the work in producing the merely attributed relation of personal identity. Contiguity, he writes, “has little or no influence in the present case”(ibid.). Presumably, he reaches that conclusion from his premise that perceptions other than those of sight and touch have no locations (*THN* 1.4.5.9-13). As for resemblance, Hume believes that it plays some role, since the ideas of memory resemble the perceptions that they copy (*THN* 1.4.6.18), but in the end he

⁴¹ Who else should he be identical to?

⁴² On Hume’s deep reliance on Locke’s treatment of identity, see Roland Hall, ‘Hume’s Use of Locke on Identity,’ *The Locke Newsletter* 5 (1974): 56-75.

concludes that “memory does not so much *produce* as *discover* personal identity, by showing us the relation of cause and effect among our different perceptions” (*THN* 1.4.7.20).⁴³

The remaining principle is causation, so Hume concludes by process of elimination, “the true idea of the human mind, is to consider it as a system of different perceptions or different existences, which are link’d together by the relation of cause and effect, and mutually produce, destroy influence, and modify each other” (*THN* 1.4.6.19). His account of personal identity as a merely attributed relation divides through his account of causation as a merely attributed relation, leaving no apparent remainder.

Hume reasons as follows. The contemplative association of ideas of cause and ideas of effect leads us from the ideas of some perceptions to the ideas of others. By his second definition of causation, when our ideas are associated in this way, and one of the perceptions precedes the other, then the earlier perception causes the later one. It follows that, in many instances, the merely attributed relation of personal identity is the merely attributed relation of causation. Perception A is part of the same person as perception B, if the idea of perception A is associated with the idea of perception B through the associative principle of cause and effect.

On this analysis, we attribute personal identity to perceptions “because of the union of their ideas in the imagination, when we reflect upon them” (*THN* 1.4.6.16). Neither causation nor personal identity is a real connection in Hume’s book, but both obtain between perceptions. He thus has the view about personal identity attributed to him in the textbooks: perceptions are bundled together into minds mostly by causal relations.

⁴³ This distinction between the production and the discovery of personal identity would make no sense if Hume were only interested in offering a diagnosis of our false belief in personal identity.

Remember, however, the relevant notion of causation is the idiosyncratic one described by his second definition.

4 Hume's Discovers the Regress

4.1 Hume's Recapitulations

We have seen how Hume generalizes his analytic methods in his chapter on personal identity. Not just causation but every putatively binding relation between two objects is to be analyzed as a merely attributed relation. No exceptions are to be made for either impressions or ideas. Given such assumptions, it seems that Hume is committed to analyzing associations of ideas as merely attributed relations. If we treat his subjectivism as sincere, a vicious regress of associations will be unavoidable.

Ironically, the nakedness of this problem has made it harder to understand Hume's acknowledgement of it. On the one hand, commentators who look at the *Treatise* through the foggy glasses of charity have had trouble understanding Hume's analyses of causation and personal identity in their intended spirit as subjectivist metaphysics. On the other hand, the obviousness of the threatened regress pushes Hume into a kind of obscuring defensiveness in his Second Thoughts on personal identity.⁴⁴ Instead of slowly and carefully explaining the difficulty, he spends most of his confession reiterating the arguments that led him to his precarious position.

After declaring his unhappiness with his own view, Hume looks back and admits, "there are two principles, which I cannot render consistent; nor is it in my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences, and that the mind*

⁴⁴ By the "Second Thoughts," I mean that part of the Appendix to Hume's *Treatise* where he describes the labyrinth into which rereading his chapter on personal identity cast him (*THN Appendix ¶¶10-21*).

never perceives any real connexion among distinct existences” (*THN Appendix ¶21*). The two problematic principles that Hume says he can’t make consistent epitomize his reasons for rejecting the two alternatives to his own theory. The first was a premise in his argument that perceptions do not inhere in an enduring substance, and the second was a premise in his argument that perceptions are not bound together by real connections. The second principle (on Humean assumptions) also generates a vicious regress in his attempt to analyze personal identity as a merely attributed relation.

Let me briefly explain how Hume uses the first principle to argue against the view that perceptions inhere in an enduring substance. I’ll then slowly explain the role of the second principle in the Second Thoughts, both as part of his restatement of an argument that perceptions are not bundled by real connections and as the assumption that undermines his own account

The first principle, recall, is “that all our distinct perceptions are distinct existences.” A little digging reveals that this is a premise in an argument that Hume has presented in Book One and presented again in the beginning of the Second Thoughts.⁴⁵ In the third paragraph of the Second Thoughts, he argues:

Whatever is distinct, is distinguishable; and whatever is distinguishable, is separable by the thought or imagination. All perceptions are distinct. They are, therefore, distinguishable, and separable, and may be conceiv’d as separately existent, and may exist separately, without any contradiction or absurdity. (*THN Appendix ¶12*)

The upshot is that perceptions don’t inhere in an enduring substance, since, if they did, they couldn’t exist apart from it.

The principle that all perceptions are distinct existences is a logical truism. It doesn’t contradict any possibly true proposition. The argument in the third paragraph shows how

⁴⁵ The Book One presentation is at *Treatise* 1.4.6.3. See Bricke’s discussion of it (*Hume’s Mind*, 67-71).

Hume gets argumentative blood out of this turnip. In effect, Hume relies on his doctrine that if $x \neq y$, then x can exist without y (*THN* 1.4.5.5).⁴⁶ If that's true, then Hume indeed has a good argument against the traditional view that perceptions inhere in a substance and can't exist without them.

Hume must intend the first principle to be a reference to his earlier argument that perceptions don't inhere in a substance. If we strip away that context, then the first principle becomes a tautology that couldn't disturb anyone. It would be a terrifically unlikely coincidence for Hume to mention the obvious fact that distinct perceptions are distinct existences twice in three pages if the passages had nothing to do with each other.

4.2 A Close Reading of the Eleventh Paragraph

Two things happen in the eleventh paragraph of the *Second Thoughts*.⁴⁷ First, Hume repeats his argument against thinking that our perceptions are bound by real connections. Second, he concedes that his policy of reducing putatively real connections into associations of idea is problematic when it comes to the association of ideas themselves.

Hume has spent the second through tenth paragraphs of the *Second Thoughts* restating his arguments that the mind isn't a substance in which perceptions inhere. He begins the eleventh by reaffirming the apparent soundness of those arguments: "So far I seem to be attended with sufficient evidence" (*THN Appendix* ¶20). In the next sentence, he implies that our perceptions must have something in common, but admits that his account doesn't succeed in saying what it is "and that nothing but the seeming evidence of the

⁴⁶ For discussions of this principle, see Bricke, *Hume's Mind*, 68-71 and Garrett, *Cognition*, Ch. 3. As I understand Hume's treatment of 'distinctions of reason' (*THN* 1.1.7.17-18), they don't constitute numerical distinctions.

⁴⁷ In Norton and Norton's enumeration, this is the twentieth paragraph of the *Appendix*.

precedent reasoning cou'd have induc'd me to receive it." This sentence contains a footnote referring to his positive account of personal identity as resting on associations of ideas.

Recall that though Hume denies that perceptions inhere in a substance, he still believes that they are somehow unified.⁴⁸ The alternative to inherence in a substratum, he writes, is to have them bound by some relation: "if perceptions are distinct existences" (and therefore do not inhere in an underlying substance), "they form a whole only by being connected together." His remaining question concerns the character of the connection.

He then restates in abbreviated form his argument from 1.4.6 that personal identity is a merely attributed relation:

But no connexions among distinct existences are ever discoverable by human understanding. We only *feel* a connexion or determination of the thought, to pass from one object to another. It follows, therefore, that the thought alone finds personal identity, when reflecting on the train of past perceptions, that compose a mind, the ideas of them are felt to be connected together, and naturally introduce each other.

The first premise is the second problematic principle. Hume again needs to assume that personal identity would be perceptible by human understanding if it were a real connection, and again he concludes that personal identity is a merely attributed relation, that is, a relation of the form, the idea of A is psychologically associated with the idea of B. The reference to merely 'feeling' a determination of thought fits with this reading, too. Hume writes that way when talking about impressions of psychological associations (e.g. at *THN* 1.3.14.20).

What comes next in the eleventh paragraph proves this interpretation. Hume expects us to be stunned and reassures us by saying that it makes his account fashionable:

However extraordinary this conclusion may seem, it need not surprize us. Most philosophers seem inclin'd to think, that personal identity *arises* from consciousness; and consciousness is nothing but a reflected thought or

⁴⁸ Melchert, 'Appendix,' 333, cf. Kail, *Realism*, 129-31.

perception. The present philosophy, therefore, has so far a promising aspect (*THN Appendix ¶120*).

The reference to most philosophers surely includes Locke in *Essay 2.27* and possibly includes Locke's friends and followers Catherine Trotter, Samuel Bold, and Anthony Collins.⁴⁹

Hume's point is that, on Lockean accounts as well as on his own, the parts of a person constitute a unity because the person is aware of them. The contrast is with a plain and realistic account, according to which a person is aware of the unity of his constituents because they antecedently form a unity.⁵⁰ Humean personal identity depends on consciousness because, according to him, personal identity is a merely attributed relation.

Just as *all our distinct perceptions are distinct existences* is a premise of Hume's argument against thinking that perceptions inhere in a substance, *the mind never perceives any real connection among distinct existences* is a premise in his argument that the mind isn't bundled together by observer-independent relations. He hasn't yet said what he thinks is wrong with his conclusion that personal identity is a merely attributed relation, and thus, he hasn't yet said why he regrets the foreclosure of rival options.

The moment of truth follows: "But all my hopes vanish, when I come to explain the principles, that unite our successive perceptions in our thought or consciousness. I cannot discover any theory, which gives me satisfaction on this head" (*THN Appendix ¶120*). Staring at this sentence out of context can lead to interpretive despair, since nothing helpful follows

⁴⁹ Trotter in *A Defense of Mr. Locke's Essay of Human Understanding*, Bold in *A Discourse concerning the Resurrection of the Same Body*, and Collins in *Reflections on Mr. Clark's Second Defence of his Letter to Mr. Dodwell and An Answer to Mr. Clark's Third Defence of His Letter to Mr. Dodwell*. I owe the references to Bold and Collins to Udo Thiel, 'Personal Identity,' in *The Cambridge History of Seventeenth Century Philosophy*, 2 volumes, edited by Daniel Garber and Michael Ayers (Cambridge: Cambridge University Press, 1998), vol. 1, 868-912 at 910n65. It's also possible that Hume intends an allusion to §89 of Leibniz's *Theodicy*. A reference in the Abstract (¶4) implies that Hume had read at least part of the *Theodicy* by 1740, as Norton and Norton observe in their editorial notes (567n4).

⁵⁰ On Kail's interpretation of Hume on personal identity (*Realism*, 131-33, 138), the causal relations that tie our perceptions together are objectively necessary and observer-independent. This is incompatible with Hume's claim to Lockean fashionability and with the structure of Hume's second argument for his theory of personal identity.

it, and, on its own, it isn't enough to determine an interpretation. There's no reason to read the sentence out of context, however. Our interpretation of the beginning of the paragraph makes it tolerably clear what Hume must mean at its end.

He doesn't say that the problem is explaining what unites our perceptions. As we have seen, he has just offered a compressed reiteration of his earlier account. Rather, Hume says that the problem is explaining what "unites our successive perceptions in our thought or consciousness." He means what unites *our ideas* of our perceptions. For Hume, thinking is what we do when we have ideas (*THN* 1.1.1), and the reference to consciousness must be read in light of the previous two sentences, which state that on his account, as on that of "most philosophers," personal identity depends on consciousness. As we've seen, his view has that consequence, because it makes personal identity depend on associations between ideas. In order to understand the passage, we need an obvious problem with the principles of association that bind our ideas of perceptions. Luckily, we have one.

If the imperceptibility of real connections between perceptions implies that personal identity is a merely attributed relation, then parity of reasoning demands that we conclude that the psychological association between ideas is also a merely attributed relation. The psychological association by which the idea of A gives rise to the idea of B can't be a merely attributed relation, however, since that would lead to an infinite regress.

Suppose that the impression that is Plato's love of mathematics is part of the same person as the impression that is Plato's love of wisdom. On Hume's account, being part of the same person is a merely attributed relation and applies to these impressions only because the *idea* that represents Plato's love of mathematics is associated with the *idea* that represents his love of wisdom. The impressions are only united indirectly "in our thought." But, Hume asks, how are we to explain the association of the corresponding ideas? By his

principles, the relevant association of ideas is not a perceptible real connection, and it must therefore be analyzed as a merely attributed relation. This entails that behind the association of the idea of Plato's love of mathematics and the idea of Plato's love of wisdom must stand another distinct association between the idea of the idea of Plato's love of mathematics and the idea of the idea of Plato's love of wisdom. And so on *ad infinitum*. But it just isn't the case that there's actually a higher-level association of ideas lurking behind every association of ideas, and if it were the case then the regress wouldn't be founded on anything solid.

The greatest exegetical dispute over Hume's metaphysics concerns his failure to recognize the tension between his psychological naturalism and his causal projectivism. That is, how can he serenely offer psychological explanations that presuppose the reality of mental causation and also offer an account of causation as a merely attributed relation? The second greatest dispute concerns the nature of his worries in the Second Thoughts. We ought to put these problems next to one another and declare victory.

4.3 The Inconsistency

A good deal of interpretive effort has centered on Hume's assertion that he can't render his two problematic premises consistent, since, obviously, the two principles don't contradict one another as they stand. How might they be supplemented by propositions that Hume explicitly or tacitly asserts to result in a genuine explicit contradiction? As I have said, the trick to seeing what he means here is to realize that he summarizes arguments by repeating a premise and assuming that his readers will remember the appended arguments.

Hume believes what is obviously true, that a person's perceptions are distinct entities that have something in common. That is,

1) A person's perceptions form a whole (Premise *THN* 1.4.2.6, 1.4.6.19, *Abstract* ¶28)

After repeating his arguments that the soul is not a substance in which perceptions inhere, Hume writes, “If perceptions are distinct existences, they form a whole only by being connected together” (*THN Appendix ¶20*) and then goes on to explain why he thought they were connected by merely attributed relations and not by real connections. These are the only three options that Hume considers:

2) If perceptions form a whole, then they either inhere in the same substance, or are connected by real connections, or are connected by merely attributed connections (Premise)

The first of Hume’s problematic principles is

3) All distinct perceptions are distinct existences (Premise *THN 1.4.6.3, Appendix ¶¶12, 21*)

In the third paragraph of the Second Thoughts, he argues

4) If x is a distinct existence, then x may exist alone (Premise *THN 1.4.6.3 Appendix ¶12*)

The argument in the third paragraph very reasonably assumes what he elsewhere states explicitly:

5) If x may exist alone, then x does not inhere in a substance (Premise *THN 1.4.5.6*)

It follows that

6) No distinct perceptions inheres in a substance (3, 4, 5)

Hume’s second problematic principle is

7) The mind never perceives any real connection among distinct existences (Premise *THN 1.4.6.16, THN Appendix ¶21*)

In 1.4.6, he assumed that

8) If there are real connections between distinct existences, then the mind sometimes perceives those connections (Premise *THN 1.4.6.16*)

It follows that

9) There aren't any real connections between distinct perceptions (3, 7, 8 *THN*

1.4.6.16)

So, it follows that

10) A person's perceptions are connected by merely attributed connections (1, 2, 6, 9 *THN* 1.4.6.16, *Appendix ¶20*)

That's the position of Book One, and of the first half of the eleventh paragraph. But a problem remains:

11) If there are any merely attributed relations between objects, then there is at least one real connection between perceptions, namely the association of ideas of those perceptions (Premise *THN Appendix ¶20*)

9, 10, and 11 can't all be true.

4.4 Some Related Interpretations

My reading of the Second Thoughts is new, so far as I can tell, but it isn't radically new.

There are many related interpretations in the literature, and comparisons may be helpful.⁵¹

Norman Melchert and Donald Ainslie agree with each other and with me that Hume's problem involves connecting our ideas of our perceptions and the looming threat of a regress. Melchert imagines three people to be analyzed and represents their perceptions of external objects with lower case letters and their ideas of those perceptions with capital letters. He asks,

What is it that makes the M's constitute Matthew, the A's Andrew and the S's Stephen? If we try to answer this question in Hume's way it is clear that only the apprehension of resemblances among them by a third-level perception

⁵¹ Good objections against many other accounts may be found in Garrett (167-80). It seems to me, as it does to Robert Fogelin (in *Hume's Skepticism in the Treatise of Human Nature*, London: Routledge & Kegan Paul, 1985, 104-05), that Garrett's own interpretation of the Second Thoughts isn't tied closely enough to the text.

will do the job. And now it is also clear that an infinite regress is under way. And it seems obvious that that cannot be right.⁵²

Similarly, Ainslie writes, “These secondary ideas, however, remain as distinct existences since there are not ideas *of them* (tertiary ideas?) associated together with the ideas of our other perceptions.”⁵³

On Melchert’s reading and on Ainslie’s, the resulting problem is that the relevant ideas of perceptions won’t be believed to stand in the personal identity relation with other perceptions in the mind. Hume, Melchert writes, “did not pretend to explain the actual identity of the self. According to the Treatise there is no such identity. What Hume had tried to do in the Treatise was to explain our belief that the self is one.”⁵⁴ On my reading, Hume’s problem is that the relevant ideas won’t stand in causal relations to one another. Both the relation of *being believed to hang together* and the relation of *causally introducing one another* provide a kind of unity between ideas, so either explaining either relation might be called a problem in explaining “the principles, that unite our successive perceptions in our thought or consciousness” (*THN Appendix 20*).

I think that Melchert and Ainslie’s interpretations get to the heart of the matter, but I don’t think that either commentator draws out consequences severe enough to match Hume’s despairing tone. The error is “considerable” (*THN Appendix ¶1*), casts him into “a labyrinth” (*THN Appendix ¶10*), and is more serious than the error of thinking that “two ideas of the same object can only be different by their different degrees of force and vivacity” (*THN Appendix ¶22*). Melchert and Ainslie’s worries just aren’t that significant. As a matter of fact, we don’t often think about the question of why we believe that our ideas of

⁵² Melchert, ‘Appendix,’ 331.

⁵³ Ainslie, ‘Reflections,’ 566.

⁵⁴ Melchert, ‘Appendix,’ 328, see also Ainslie, ‘Reflections,’ 563-66.

our perceptions are parts of the same person as the rest of our perceptions, and, if we did think about this, then, yes, we could form tertiary ideas and associate them in the imagination.

Indeed, on Ainslie's interpretation, the problem is even more restricted than that. He believes that since "most people do not ever consider their minds—or only rarely and indeterminately,"⁵⁵ Hume's question of the origin of the belief in the simplicity and identity of the mind is only a question about the psychology of certain philosophers. I think that since Hume's emphasis is on his own "contradictions" (*THN Appendix ¶10*) and not the mistakes of others, he can hardly be obsessing over the question of why some philosophers mistakenly think that their ideas of their perceptions belong to their minds.

As we've just seen, Hume describes his difficulty as that of describing the principle that ties "all our particular perceptions" together. On Melchert's and Ainslie's accounts, he would have succeeded explaining how the mind forms an idea of the self that contains almost all our perceptions together, except for a few stragglers which could be picked up with a bit of reflection. Nothing important in Hume's system really hangs on whether the ideas that we form of our minds are themselves always believed to be parts of the same mind. On his view, "All the disputes concerning the identity of connected objects are merely verbal, except so far as the relation of parts gives rise to some fiction or imaginary principle of union" (*THN* 1.4.7.21). If there is no unification of our associating ideas in the imagination, then it follows by Hume's lights that the question of whether those ideas are part of the self is only terminological.

If association is a causal relation, and causal relations are merely attributed relations, then the reasons that Melchert and Ainslie give for thinking that Hume can't explain why

⁵⁵ Ainslie, 'Reflections,' 563.

those secondary ideas are believed to be part of the same person would also be reasons for thinking that those secondary ideas can't introduce one another. If his second definition of causation requires such introductions behind every causal connection between perceptions, then his system would grind to a halt. Without associations between secondary ideas, our perceptions won't stand in causal relations to one another. That conclusion would contradict almost every section of the *Treatise*.

My reading is quite similar to Vijay Mascarenhas's. He writes,

Remember that Hume denied the existence of "real" relations by resolving them into "ideal" relations: there is no real necessary connection, for example, between the impressions of cause and effect, only an 'ideal' or associative one between the *ideas* of those impressions. . . . When it comes to the belief in personal identity, however, Hume reaches a dead end in the labyrinth, for his explanatory apparatus wraps around itself: the ideal relations into which real relations are resolved themselves rely on something that not only explains the formation of the belief in personal identity, but would amount to, if not a constant and invariable self, then at least some kind of unity of consciousness.⁵⁶

We agree that the association of ideas presupposes something that Hume thinks can't be provided. I think that the problem is that the association of ideas is a causal relation, and thus requires an endless hierarchy of higher-level ideas. Mascarenhas's primary worry is that association presupposes the unity of consciousness: "my focus is on whether the mind is unified enough to account for the association of ideas."⁵⁷

I do not see, however, that Hume believes that the association of ideas presupposes the unity of consciousness. Setting aside the secondary role that resemblance plays, Hume's theory of personal identity comes down to saying that the composition of the self depends on causal connections between perceptions. That commits him to saying that causal

⁵⁶ Vijay Mascarenhas, 'Hume's Recantation Revisited,' *Hume Studies* 27 (2001): 279-300, at 295

⁵⁷ *ibid.*, 300n34,

connections between perceptions don't in turn depend on the composition of the self,⁵⁸ but Mascarenhas hasn't cited texts that show that Hume was unhappy with this position. On Mascarenhas's interpretation, Hume changed his mind about whether association of ideas was prior to the composition of the mind. Such a change of heart would call for a relatively straightforward concession and recantation. I don't think that Hume would have described himself as being cast in a labyrinth or rehearsed all the subtleties of the Second Thoughts if that were all that was going on.

John Haugeland offers a more pointed version of this worry. His interpretation, like mine, rests on the difficulty in combining Hume's theory of causation with his theory of personal identity. Unfortunately, rather than working with Hume's second definition of causation, Haugeland appeals to his first, which doesn't fit the text of the Second Thoughts very well. He writes, "In a nutshell, the question of personal identity is how we can allocate all the conceivably free-floating perceptions into various personal bundles, given that the prerequisite pattern of constant conjunctions constitutive of mental causation presupposes a *prior* bundling."⁵⁹ Haugeland thinks that Hume can't solve this problem, since he "wants the coherent patterns (mental causation) to do the whole job. Unfortunately, there are no coherent patterns unless the perceptions are already sorted—all he has is perception salad."⁶⁰

According to Haugeland, without previously established patterns of constant conjunction, Hume can't establish any of the causal connections needed to bundle perceptions together as a whole. As I see it, Hume is willing to give himself previous patterns of constant conjunction, and his immediate problem concerns the absence of

⁵⁸ "The same principle cannot be both the cause and effect of another; and this is, perhaps, the only proposition concerning that relation which is either intuitively or demonstratively certain" (*THN* 1.3.6.7).

⁵⁹ John Haugeland, 'Hume on Personal Identity,' in *Having Thought: Essays in the Metaphysics of Thought*. (Cambridge, Mass.: Harvard University Press, 1998), 63-71, at 68.

⁶⁰ *ibid.*, 69.

associative mechanisms behind the associations between our ideas of our perceptions. I don't think that Haugeland's reading can make sense of the divided structure of the crucial eleventh paragraph. Hume does mention the "train of past perceptions," in that paragraph, but only in the first, more optimistic part, where they seem to be unproblematic:

the thought alone finds personal identity, when reflecting on the train of past perceptions, that compose a mind, the ideas of them are felt to be connected together, and naturally introduce each other. However extraordinary this conclusion may seem, it need not surprize us. (*THN Appendix ¶20*)

The difficulty only arises in accounting for union of our ideas of those perceptions. The first part of the paragraph, I think, sends us upward, looking for ideas of perceptions. Hume wants his readers to look upward again, for ideas of those ideas, and to realize that the ideas will give out, sooner rather than later.

A successful interpretation of the Second Thoughts should meet at least the following four desiderata. First, it should raise a serious enough difficulty to match Hume's despairing tone. Second, it should be more than a mere change of mind, lest his elaborate presentation of his problem and his description of it as a labyrinth be utterly incongruous. Third, it should succeed in offering a close, accurate reading of the eleventh paragraph, where Hume comes closest to explicitly stating the problem. Fourth, it should explain the contradiction that Hume sees between his two problematic principles.⁶¹ As far as I know, my interpretation is the only one that meets all these requirements.

4.5 Hume's Description of the Problem

If the problem of the Appendix boils down to the worry that Hume's second definition of causation generates a vicious regress, we might well wonder why he discusses it in the context of his account of personal identity. As a matter of fact, he doesn't say that the

⁶¹ Ainslie, 'Reflections,' 569-70.

difficulty he examines is essentially about personal identity. He just writes that he found it upon rereading that chapter. “Upon a more strict review of the section concerning *personal identity*,” Hume reports, “I find myself involv’d in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent” (*THN Appendix* ¶20). The fact that he found the regress by rereading 1.4.6 is compatible with the fact that he could have found the regress by rereading 1.3.14. If he had, of course, the Second Thoughts would have been differently presented. We don’t need an explanation of why he found the regress in one passage and not in another.

A bit of speculation may be in order, anyway. Let me offer one hypothesis founded on a detail and another more general hypothesis. The detail is that Hume had addressed the issue in *Treatise* 1.3.14 with his “subtlety, which it will not be difficult to comprehend” (*THN* 1.3.14.28). We have observed that his remarks are not enough to solve the problem, but presumably they were enough to make him think that he had handled it. The more general suggestion is that the discussion in the chapter on personal identity highlighted the problem in Hume’s second definition. In 1.3.14, Hume argues that efficacy isn’t in the world, but rather “belongs entirely to the soul” (*THN* 1.3.14.24). In 1.4.6, Hume argues that, with respect to his theory of real connections, the soul is part of the world. Finally, in the Appendix, he realizes that the two positions are inconsistent.

5 Hume’s Objective Turn

5.1 Revisions in the *Enquiry*

Ainslie writes, “Perhaps the most difficult problem for most interpretations of the ‘Appendix’ is that they fail to explain why Hume thinks that his discussion of personal

identity contains his single ‘very considerable mistake’ (T.623) in all Books I and II.”⁶² At first look, my reading succeeds triumphantly here. Hume thinks that the tension between his psychological naturalism and his causal projectivism is the greatest mistake in Books 1 and 2 of the *Treatise*, because it is, in fact, the greatest mistake in Books 1 and 2 of the *Treatise*.

This is not, however, what Ainslie means. According to him, Hume “seems to find the problem he diagnoses in the ‘Appendix’ to be a limited one, affecting only the Section and not any other part of his treatment of the understanding.”⁶³ I disagree, but Ainslie has his reasons, writing

most of the rest of Book I (the discussion of space and time, Part ii of Book I, is the notable exception) re-appears in one form or another in the first *Enquiry*, which he describes as differing from the *Treatise* only in the “manner” in which his views are presented.⁶⁴

That isn’t exactly what Hume says in the passage that Ainslie cites. Rather, Hume claims that the *Treatise’s* poor sales “had proceeded more from manner rather than the matter” (LDH i.3=DNH 5). ‘Matter’ might mean *topic* in this context, but even if it means *content* surely Hume’s diagnosis of poor sales must be compatible with at least some changes. The *Enquiry* is a good deal shorter than Book 1 of the *Treatise*, after all. More importantly and more dramatically, in a notice written right before his death and prefixed to an edition of the *Enquiries*, Hume complained that people who criticized him by quoting the *Treatise* engaged in a “a practice very contrary to all rules of candour and fair-dealing” and declared, “the Author desires, that the following Pieces may alone be regarded as containing his philosophical sentiments and principles” (*EHU Advertisement*).⁶⁵ This is a more forceful repudiation of earlier views than anything we find in the published writings of Plato or

⁶² *ibid.*, 574.

⁶³ *ibid.*, 574-75.

⁶⁴ *ibid.*, 575, see also Melchert, ‘Appendix,’ 331.

⁶⁵ See Kemp Smith, *Philosophy*, 530-36.

Wittgenstein, to name two philosophers whose work almost all commentators are willing to divide between early and late.

In defense of his assertion that Hume doesn't seriously revise his views on topics outside the chapter on personal identity, Ainslie also quotes from a letter to Gilbert Eliot where Hume writes of the *Treatise* and the *Enquiry*, "The philosophical Principles are the same in both."⁶⁶ In the same letter, however, Hume writes of the *Treatise*, "So vast an Undertaking, plan'd before I was one and twenty, & compos'd before twenty five, must necessarily be very defective. I have repented my Haste a hundred & a hundred times" (*LDH* i.158). Given Hume's talk of youth, defect, and repentance, the letter can hardly support the view that there are no doctrinal differences between the two works. The assertion that the principles are the same and that he repented his haste two hundred times are obviously compatible: the general principles are the same, but he made serious changes in his applications of those principles. In the prefixed notice, he writes, "in the following pieces, . . . some negligences in his former reasoning and more in the expression, are, he hopes, corrected" (*EHU Advertisement*). Whatever negligent expressions Hume had in mind, at least some of the improvements in the *Enquiries* were corrections in reasoning.

In any case, Hume isn't some pre-Socratic philosopher for whom we have only doxographic reports. We can just read the texts of the *Treatise* and the first *Enquiry* and look for differences. Let me list five.⁶⁷

First, in the *Treatise* Hume's second definition of cause appeals to both a contemplative association and a lively one: "A CAUSE is an object precedent and contiguous

⁶⁶ Ainslie, 'Reflections,' 575n28.

⁶⁷ Eric Schliesser has a good paper ('Two Definitions of Causation, Normativity, and Hume's Debate with Newton,' in *Newton in Historical Context*, edited by Steffen Ducheyne, Brussels: Royal Flemish Academy of Sciences, forthcoming) where he draws ten distinctions between the definitions of causation as they stand in the *Treatise* and the *Enquiry*. My first and fourth differences correspond to his second and first, and my third difference is his fourth.

to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other” (*THN* 1.3.14.31 170/114). In the *Enquiry* the definition is simplified to a single association between impression and idea, and a cause becomes “an object followed by another, and whose appearance always conveys the thought to that other” (*EHU* 7.29).

Second, in the *Treatise* Hume concludes, “The efficacy or energy of causes is neither plac’d in the causes themselves, nor in the deity, nor in the concurrence of these two principles; but belongs entirely to the soul” (*THN* 1.3.14.23) and he explicitly and approvingly cites Malebranche’s arguments against traditional accounts according to which bodies are second causes (*THN* 1.3.14.7n29). In the *Enquiry*, Hume mocks “modern metaphysicians” who “rob second causes of all force or energy” and who follow Malebranche in denying “that matter has a real, though subordinate and derived power” (*EHU* 7.25n16).

Third, in the *Treatise*, Hume’s first definition makes us consider causation as a philosophical relation and the second makes us consider it as a natural relation (*THN* 1.3.14.31). In the *Enquiry*, the distinction between natural and philosophical relations is omitted.

Fourth, in the *Treatise*, Hume’s first definition of causation isn’t supplemented by a modal paraphrase. In the *Enquiry*, after Hume defines “a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second.*” He then paraphrases the definition as follows: “Or in other words, *where, if the first object had not been, the second never had existed*” (*EHU* 7.29). Fifth, there is no discussion of personal identity in the *Enquiry Concerning Human Understanding*.

5.2 The Differences Explained

I now explain these changes. In the *Enquiry*, Hume moved away from subjectivism in metaphysics by consistently hewing to the idealized interpretation of his second definition of causation. As we have seen, on the idealized version of the second definition no regress arises. By *stipulation*, the idealized observer associates the ideas of a cause and the idea of the effect if the things resembling the cause are constantly conjoined to things resembling the effect. This stipulated, hypothetical association of ideas doesn't require an actual association of ideas to back it up, and thus the regress is snipped in its hypothetical bud.

Hume's appeal to hypothetical observers explains the change of form in the second definition from the *Treatise* to the *Enquiry*. In the *Treatise*, in his easy-to-understand subtlety, he attempted to account for the lively association between impression and lively idea by appealing to an underlying contemplative association between idea and idea. As a result, he bifurcated his second definition of causation. By the time he wrote the *Enquiry*, he came to see that this half-measure does not succeed in giving a complete account of causal connections in the mind. By consistently adopting an idealized interpretation of the observer in the second definition, he solved the problem in a different way, and thus could simplify the definition.

This flight from subjectivism likewise explains Hume's shifting attitude to occasionalism. In the *Treatise*, Hume thought that bodies weren't first-rate examples of causes, but rather genuinely depended on actual associations of ideas. In this context, he was willing to make common cause with Malebranche. In the *Enquiry*, Hume changed his mind and concluded that causal connections between bodies are as good as causal connections between any two entities. Inertial force should be analyzed by constant conjunction. Reference to 'vis inertiae' denotes the facts of experience "that a body at rest

or in motion continues for ever in its present state, till put from it by some new cause; and that a body impelled takes as much motion from the impelling body as it acquires itself” (EHU 7.25n16). ‘Gravity’ is to be given a similar interpretation. On this understanding of physical forces, Malebranche’s denial that bodies are causes becomes an object of derision, suitable for foreigners, but one which, in previous generations, “had, however, no authority in England” (ibid.). The theory of causation in the *Enquiry*, unlike that of the *Treatise*, no longer maps neatly onto the theory of *The Search After Truth*. Hume has given up on the doctrine that the soul is the hidden engine of causation.

The distinction between natural and philosophical relations is omitted in the *Enquiry*, because the second definition has lost its superior status. In the *Treatise*, the second definition no longer captures a true relation while the first definition describes a relation in the philosopher’s broad sense. The two definitions are equivalent in *Enquiry* and stand on the same footing.

With respect to the fourth change, Antony Flew has argued that Hume’s gloss on the first definition (“where, if the first object had not been, the second never had existed”) cannot be taken as a genuine equivalent of the definition, since “subjunctive conditionals cannot validly be deduced from statements of mere conjunction.”⁶⁸ Flew is making an evaluative point, which I don’t want to dispute, but I do want to add an interpretive point. Suppose that the first definition in the *Enquiry* is a full-fledged definition of causation, one that Hume believes captures the necessary connection between cause and effect. Such necessities are discovered by experience and not by reason, but he intends to offer us a genuine account of necessity. If so, then Hume has indirectly offered an analysis of a central

⁶⁸ *Belief*, 131; see also David Lewis, ‘Causation’ with postscripts, in *Philosophical Papers*, vol. 2. (Oxford: Oxford University Press, 1986), 159-213, at 159-61.

concept of modality, and the analysis should be expected to sustain the counterfactuals entailed by causal claims, including *if the cause hadn't occurred then the effect wouldn't have occurred*. That is to say, if *A causes B* may be properly analyzed by declaring that A is “an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second” then *B wouldn't have existed without A* may be properly analyzed in the same way.⁶⁹ The modal paraphrase is a genuine paraphrase because Hume takes the initial analysis to be a genuine and full analysis of causation.

If this interpretation of Hume's paraphrase of the first definition is correct, then the first definition has lost the second-class status that it had in parts of the *Treatise*. Recall that in his reply to the shocked interlocutor, Hume had said that contiguity, precedence, and the repetitions of those relations are mind-independent, but that “if we go any farther, and ascribe a power or necessary connexion to these objects; this is what we can never observe in them, but must draw the idea of it from what we feel internally in contemplating it” (*THN* 1.3.14.28). In the *Treatise*, Hume tells us that his first definition of causation doesn't go far enough for us to ascribe a necessary connection between the objects. Hume implies in the *Enquiry* that the first definition supports counterfactuals, which suggests that he no longer believes that the first definition is incomplete with respect to necessity. This would make sense on the idealized interpretation of the second definition, since the second definition reduces to the first on that interpretation.

For our purposes, the most important difference between the *Treatise* and the *Enquiry* is that the *Enquiry* lacks the discussion of personal identity. This excision is to be expected on my interpretation. Recall the argument that Hume offers at the end of 1.4.6 for his

⁶⁹ See Flew, *Belief*, 131-32. I set aside certain borderline cases such as causal overdetermination. For a discussion of some of these subtleties, see Lewis, ‘Causation,’ 170-213.

account of personal identity. Because we can't perceive any real bonds between perceptions, the connections between perceptions that bind them together to make a self must depend on associations of ideas. There are three possible principles of association that could do the trick: those of contiguity, resemblance, and causation. Hume eliminates contiguity and resemblance, and what remains is that two perceptions are part of the same person if the ideas of those perceptions are bound together by the associative principle of cause and effect.

With the shift from mostly subjective to entirely idealized readings of the second definition, associations of ideas no longer do metaphysical work in the *Enquiry's* account of causation. Hume no longer believes that the imperceptibility of a putatively real connection entails that the relation depends on an association of ideas. So we would expect him to lose faith in his argument for his account of personal identity.

Ainslie notes that the “bundle view . . . re-appears in Hume's final work, the *Dialogues Concerning Natural Religion*” (DNH 61).⁷⁰ Hume may have retained the view that our perceptions are bundled together; my point is he had lost any faith that he had provided a proper account of *how* they were bundled together. Evidence of this loss of confidence comes six years after the publication of the Appendix and two years before the *Enquiry* in a letter to his cousin, the eventual Lord Kames. Hume writes, “I like exceedingly your Method of explaining personal Identity as more satisfactory than any thing that had ever occur'd to me” (NLDH 20).⁷¹

⁷⁰ Ainslie, ‘Reflections,’ 577n30.

⁷¹ Albert Tsugawa. ‘David Hume and Lord Kames on Personal Identity,’ *Journal of the History of Ideas* 22 (1961): 398-403 has a useful discussion of the issues raised by the letter.

5.3 The Regress Resolved

We may take one of two approaches to seeing how Hume has resolved the contradiction that I formulated in §4.3. If we take ‘merely attributed relations’ to extend to relations between A and B where the idea of A is only hypothetically associated with the idea of the idea of B, then, by the time of the *Enquiry*, Hume would reject 11, the premise that if there are any merely attributed relations between objects, then there is at least one real connection between perceptions. On the idealized interpretation, it no longer follows from the fact that there are merely attributed relations that there must be at least one real connection.

If we take ‘merely attributed relations’ more narrowly, so that they extend only to relations between A and B where the idea of A must be actually associated with the idea of B in some existing observer, then he would reject 2 (“If perceptions form a whole, then they either inhere in the same substance, or are connected by real connections, or are connected by merely attributed connections”) since there’s now at least one other option. So, for example, all putatively real connections might be causal connections which are not merely attributed relations in the restricted sense, but nevertheless depend on covering laws and “circumstances foreign to the cause” (*EHU* 7.29).

In the *Enquiry*, Hume does not say or imply that causal connections depend on our causal inferences. The doctrine that efficacy is in the mind does live on in two passages in the *Enquiry*, but only in a domesticated form.

The first passage occurs right after Hume has stated his account of the origin of our idea of connection. He writes, “when we say, therefore, that one object is connected with another, we mean only, that they have acquired a connexion in our thought, and give rise to

this inference, by which they become proofs of each other's existence" (*EHU* 7.28).⁷² Hume is part of an early modern tradition in semantics according to which assertions 'mean' the inner states they indicate.⁷³ In the *Enquiry*, as in the *Treatise*, to investigate the meaning of expressions is to investigate "the nature and origin of those ideas, we annex to them" (*THN* 1.3.14.14). Thus, Hume's explication of what we mean when we say that two things are connected tells us what he thinks goes through the mind of someone who sincerely asserts that one thing causes another. It doesn't tell us the truth conditions of that thought. Which is to say, it doesn't tell us how to apply the second definition. For that, we need to turn to a footnote in section eight of the *Enquiry*.

The footnote proves that Garrett's absolute interpretation of Hume's second definition is not just a clever move by a commentator to get his philosopher out of a jam, but precisely describes Hume's considered judgment of how the second definition is to be applied. Since it contains the *Enquiry's* clearest mention of the doctrine that necessity is in the mind, and it coheres perfectly well with the rest of the work, the passage may serve as his last word on the matter.

In the footnote, Hume offers a diagnosis of the fact that people believe that their acts are undetermined. He begins with a new formulation of the doctrine that efficacy is in the mind:

The necessity of any action, whether of matter or of mind, is not, properly speaking, a quality in the agent, but in any thinking or intelligent being, who may consider the action; and it consists chiefly in the determination of his thoughts to infer the existence of that action from some preceding objects (*EHU* 8.22n18).

⁷² A referee pointed this passage out to me.

⁷³ Walter Ott, 'Hume on Meaning,' *Hume Studies* 32 (2006): 233-52.

Hume's distinction between observer and agent and the characterization of the observer as one who "may consider" the action imply that a merely hypothetical observation suffices for necessity.

The rest of the passage confirms this reading. Hume observes that when performing actions, we sometimes feel something like "a looseness, or indifference," but when reflecting on them, we do not, "but are commonly able to infer them with considerable certainty from their motives" (ibid.). Even if we imagine that what we do is undetermined,

a spectator can commonly infer our actions from our motives and character; and even where he cannot, he concludes in general, that he might, were he perfectly acquainted with every circumstance of our situation and temper, and the most secret springs of our complexion and disposition. Now this is the very essence of necessity, according to the foregoing doctrine (ibid.).

For Hume's purposes, the important point is that agents are not authoritative judges of whether their actions are determined. Rather, what matters is the judgment of a well-informed observer. For our purposes, the important point is that Hume explains his formula that necessity is in the mind by appealing not to any actual observer but to one hypothetically armed with full knowledge of the hidden mental qualities of the agent. In this way, Hume resolves his regress.