COMMON FICTIONS AND HUME’S DILEMMA ABOUT THE SELF

Matthew Parrott

In the Treatise on Human Nature, Hume characterizes many of our most fundamental thoughts as "fictions". These include, for example, our beliefs about enduring material objects in the external world, persisting selves, and absolute time. Unlike the kinds of fictions we might find in a fairy tale or dream, these fictions exist in every human mind; they seem to be universal features of human cognition. I shall therefore refer to them as common fictions.

Much of Book I of the Treatise is devoted to explaining the existence of these common fictions. Our thoughts about necessary connections, enduring material objects, empty space and persisting selves are each characterized by Hume as fictitious. Although he offers a distinct explanation for each one of these, I believe they all share the same basic structure. Hume thinks that the mind generates a common fiction because it plays a unique and important functional role. Specifically, the function of a common fiction is to relieve a kind of psychological tension that can naturally arise between how a certain psychological process or transition feels and how it is understood to be. In order to eliminate this tension, Hume believes the mind will "feign" a common fiction. Thus, for Hume, the functional role common fictions have in our psychological lives explains their existence.¹

Why do psychological tensions arise and how precisely does the existence of a common fiction alleviate them? For Hume, it is not a conceptual truth about the nature of a common fiction that it is feigned in order to relieve psychological conflict. Common fictions are not defined or individuated by their functional role. Rather, it is a matter of empirical

¹ Garrett, Don. 2006. "Hume's Naturalistic Theory of Representation". Synthese 152.3: 301-319 also discusses functionalism in Hume. He argues that "all representation consists, for Hume, of one thing playing, by means of the mental effects and mental dispositions it produces in particular circumstances, a significant part of the causal and/or functional role of what it represents." (pg. 310) In this paper, I am primarily concerned with how functional roles explain the existence of common fictions, not with the representational content of ideas. Nevertheless, it will also seems true in some cases that what a fiction is supposed to represent should, as Garrett argues, also be understood in terms of its functional role.
fact that our minds generate common fictions for the purpose of relieving psychological tension. If Hume believes this, it is because he thinks the evidence from particular cases supports it. Thus, the best way to demonstrate that Hume’s various explanations common fictions do share the same general functional structure is to look at the details of his specific explanations. However, because Hume presents a unique explanation for each common fiction, it would take up too much space to discuss them all within a single essay. Instead, I shall focus on just one, perhaps his most famous.²

Hume’s well-known account of the common fiction of a persisting self offers perhaps the best example. The primary reason for this is that, by Hume’s own lights, the explanation ultimately fails. In the Appendix he reports a problem with his positive account of personal identity so formidable that it leads him to abandon his original view. But he never offers a very clear or precise statement of precisely what this problem is. In this paper, I shall argue that Hume’s positive account fails because, as he comes to realize in the Appendix, the common fiction of a persisting self cannot be functionally explained in the same manner as other common fictions. Because Hume has no other way of accounting for its existence, he retracts his positive account and pleads the "privilege of the skeptic".

Hume’s sudden retraction in the Appendix is a familiar topic and, over the years, many different interpretations of this brief passage have been presented. Of the existing

2 But what sort of thing is a common fiction? The ontological status of fictions is an interesting issue and it has received very little attention. In one of the few places where this question is addressed, "Impressions, Ideas, and Fictions," Hume Studies 13.2 (1987), 381-389, Saul Traiger suggests that fictions are ideas "applied to something from which they cannot be derived." (1987, pg. 397) My own view is that Hume describes two distinct kinds as fictions: both peculiar imaginary ideas like a golden mountain and certain beliefs count as fictions. Both are fictional in the sense that what they represent is not real. Thus, the idea of a golden mountain is a fiction because there are no golden mountains. Similarly, in 1.2.5 Hume describes time without change as "fictitious duration". Because the idea of time just is the idea of a succession of changing perceptions, when we believe that time exists without change, we represent something that is not real, hence "fictitious duration". Cf. Anthony Brueckner, “Humean Fictions” Philosophy and Phenomenological Research. 1986. 464.
ones, more than one strikes me as plausible.\footnote{In particular, Donald Ainslie. 2001. "Hume's Reflections on the Identity and Simplicity of Mind," Philosophy and Phenomenological Research 62.3: 557-578; 2008. "Hume on Personal Identity," in A Companion to Hume; Donald Baxter. 2009. Hume's Difficulty: Time and Identity in the Treatise; and Abraham Roth. 2000. "What Was Hume's Problem with Personal Identity?" Philosophy and Phenomenological Research 61.1 all strike me as plausible readings of Hume's Appendix.} I think this is largely because the text of the Appendix is vague enough to support many different readings. For this reason, I also believe it is extremely difficult, if not impossible, to conclusively settle the historical question of what exactly Hume believed was problematic with his account of personal identity. Nevertheless, I shall offer yet another reading of the Appendix and situate my own interpretation among the existing alternatives. The primary reason to favor the reading presented in this essay is that it illuminates a common element between Hume's account of the fiction of the self and the other common fictions he explains in the Treatise. Hume's understanding of why we believe in a persisting self is the same type of explanation he offers for many thoughts. Yet, this central aspect of his work has been neglected by many of his readers. It is thus frequently said that, when compared to Kant, Hume is globally obscure. My hope is that this essay sheds light on this alleged obscurity.

I

When Hume first mentions the topic of personal identity in the Treatise he warns us that "there is no question in philosophy more abstruse than that concerning identity, and the nature of the uniting principle, which constitutes a person." (T 1.4.2.5; SBN 189) Hume immediately recognizes there is no impression or experience from which we can directly copy an idea of a persisting unified self; nothing of this sort originates in experience. Even when we self-consciously turn our attention inward toward our own psychological life and
try to observe the self, we can grasp only an individual perception, a particular thought, sensation, or feeling. As Hume famously puts it:

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. I never catch myself at any time without a perception, and never can observe any thing but the perception. (T 1.4.6.3; SBN 252)

Hume also has given us other arguments (T 1.4.2; T 1.4.5) for thinking that perceptions are the only things that we are ever immediately aware of and he believes that "since all our perceptions are different from each other, and from everything else in the universe, they are also distinct and separable and may be consider'd as separately existent, and may exist separately, and have no need of anything else to support their existence." (T 1.4.5.5; SBN 233) This last conclusion is reiterated at the very beginning of the section "Of Personal Identity". But, if all thought is nothing but a succession of distinct and distinguishable perceptions, where does our notion of the self come from?

In his attempt to answer this last question, Hume actually sets out to accomplish two separate tasks that are frequently not distinguished. First, he analyzes what he characterizes as our "true idea" of the self. According to Hume, our idea of the self is really an idea of "a heap or collection of different perceptions" (T 1.4.2.39; SBN 207). No real connection unites any of these distinct perceptions and there is therefore no unified or simple self. As Hume says "they are the successive perceptions only, that constitute the mind." (T 1.4.6.4; SB 253) To the extent that our talk about the self is ever meaningful, it is because we refer to this bundle of perceptions; it is our "true idea" of the self. Nevertheless, we also ordinarily think

---

4 "All these are different, and distinguishable, and separable from each other, and may be separately consider'd, and may exist separately, and have no need of any thing to support their existence." (T 1.4.6.3; SBN 252)

5 One reason this is important for him is because the idea of the self seems to figure prominently in Book I of the Treatise. If there were no "true idea" of the self, it would undermine much of what Hume does in Books II and III because there would be no meaningful way for us to talk about persons or selves. This is the biggest
of this bundle of perceptions as an entity that persists through time. So we also need to understand why we think the self is "possest of an invariable and uninterrupted existence". (T. 1.4.6.5; SB 253)

This second task is more challenging. Because the true self consists of only distinct "successive perceptions", Hume claims "the identity, which we ascribe to the mind of man, is only a fictitious one." (T 1.4.6.15; SBN 259) To ascribe fictitious identity to the self just is to believe that it persists through time. Since each of us believes this, the fiction of a persisting self is a common fiction. Because Hume has used his principles of association to explain the existence of other common fictions, like our belief in necessary connection, he is confident that the explanation of our belief in a persisting self "cannot, therefore, have a different origin, but must proceed from a like operation of the imagination upon like objects." (T 1.4.6.15; SBN 259) The "operation of the imagination" to which Hume refers is its disposition to "transition" between ideas related in accord with the principles of association. In the case of the self, a series of ideas that resemble one another or are causally related to one another will "facilitate[s] the transition of the mind from one object to another, and render[s] its passage as smooth as if it contemplated one continued object." (T 1.4.6.6; SBN 254) A series of perceptions related by causation or resemblance is so remarkably similar to one single and uninterrupted thing that Hume thinks we make a certain type of mistake. He thinks the similarity "makes us substitute the notion of identity, instead of that of related objects." (T 1.4.6.6; SBN 254) Considering an example should help us see

difficulty facing Norman Kemp Smith's interpretation of Hume's Appendix in The Philosophy of David Hume (London: Macmillan, 1941). Kemp Smith alleges that because Hume denies that we have any impression of the self he thinks that we do not have any idea of the self. This leads to an inconsistency, which Kemp Smith argues Hume eventually recognizes, with the discussion of self-awareness in Book II. This interpretation overlooks Hume's insistence in Book I that there is a "true idea" of the self that, like any complex idea, is copied, albeit indirectly, from impressions. For a more detailed discussion of Kemp Smith's interpretation, see Don Garrett, "Hume's Self Doubts about Personal Identity," The Philosophical Review XC.3 (1981): 337-58.
more clearly what Hume means. I shall intentionally oversimplify the example in order to make a few key points more clear.⁶

Suppose that I observe an oak tree. If I perceive the tree uninterruptedly between 11:00 am and 11:05 am, my perception of the tree is "uninterrupted and invariable" and it feels like I am contemplating a single unchanging object over time. If I were to reflect on my perceptions of the oak at the different times, for example if I were to distinguish the 11:00 perception from the 11:05 perception, I would naturally say that they are both of the same oak. In Hume's words, we "suppose the change to lie only in the time, and never exert ourselves to produce any new image or idea of the object."⁷ (T 1.4.2.33; SBN 203) But, now suppose that I turn my attention away from the tree toward something else and then, at 11:10, return to observing the tree. My perceptions of the oak are now interrupted and this interruption makes vividly clear that the perception of the oak at 11:05 and the one at 11:10 are not identical; they are not the same thing. They do however resemble each other very closely and Hume believes that "the smooth passage of the imagination along the ideas of resembling perceptions makes us ascribe to them a perfect identity." (T 1.4.2.36; SBN 205) He thinks that the interrupted psychological progression from the 11:05 oak-perception to the 11:10 oak-perception is so smooth and easy that it feels just like contemplating a single unchanging thing from 11:00 to 11:05.

The fact that these two distinct psychological processes, the one uninterrupted and invariable and the other interrupted and variable, feel the exact same way explains why I

---

⁶ In particular I am focusing on a much smaller number of distinct perceptions that Hume believes would be necessary to generate a belief in a distinct body. See T 1.4.2.42.

⁷ Strictly speaking, since this oak perception does not change, the idea of time does not apply. Thus, the "supposition" of a change in time is itself a fiction. Without this fiction, we would not be able to even acquire the idea of identity in the first place, which is an idea between unity and number. For more on this, see Baxter (2009).
think there is only one persisting oak tree in the case where there is interruption. I confuse
the two processes because of how they feel. Hume describes this as follows:
An easy transition or passage of the imagination along the ideas of these different and
interrupted perceptions, is almost the same disposition of mind with that in which we
consider one constant an uninterrupted perception. This therefore very natural for us to
mistake the one for the other. (T. 1.4.2.35; SBN 204)

The result of confusing these processes is a belief that there is a single persisting oak tree, a
"continu'd being, which may fill those intervals, and preserve a perfect and entire identity to
our perceptions". This belief in an oak tree as a distinct object that continues to exist
independently of our perceiving it is, on Hume's view, a common fiction. (T. 1.4.2.36;
SBN 205)

The belief in a persisting self is a common fiction in a similar way. The psychological
process along a series of closely related ideas seems like the process of contemplating a
single unchanging thing and this makes us believe there is a persisting self despite the
interruptions and changes in our perceptions. As in the oak tree example, in the case of the
self, Hume thinks that a "progress of thought" is produced by a succession of perceptions
related by either resemblance or causation. But unlike the case of the oak tree, in this case
the progression is along a "train of ideas" instead of along a series that includes impressions.
Hume calls perceptions of other perceptions "secondary ideas" and we believe in a persisting
self because our imagination progresses along a series of secondary ideas related by
resemblance and causation. This is why Hume believes that memory is necessary for having

---

8 According to Hume, a series of resembling perceptions will naturally "convey the imagination more easily
from one link to another, and make the whole seem like the continuance of one object." (T 1.4.6.18; SBN 261)
Causation is slightly more significant in his account of the self because Hume thinks the perceptions that make
up a self are all "linked together by the relation of cause and effect, and mutually produce, destroy, influence,
and modify each other." (T 1.4.6.19; SBN 261) The causal relation also helps Hume make sense of how a
person can be identical to their four-year-old self without resembling her in any way.

9 This point is heavily emphasized by Ainslie in his (2001) and (2008). Ainslie rightly insists that "For Hume's
account of imperfect identity to apply to minds, the association of ideas that this account involves must be the
thoughts about a self that persists through time; it "alone acquaints us with the continuance and extent of this succession of perceptions." (T 1.4.6.20; SBN 261) Memories are perceptions of other perceptions or secondary ideas. Thus, in memory the imagination progresses along a series of closely related secondary ideas. Because this feels just like contemplating something that is invariable and uninterrupted, Hume thinks we attribute identity to the changing series of perceptions and thereby come to believe in a persisting self.

II

In the Appendix, Hume surprisingly retracts this positive account of our belief in a persisting self. What problem did he recognize? The answer to this question has proven to be elusive. This is because Hume's statement of his reasons for retracting his account occurs in two very obscure paragraphs. He begins the first by claiming he encounters a problem with his attempt to "explain the principle of connexion" which binds a person's perceptions together and "makes us attribute to them a real simplicity and identity." (App 20; SBN 635) Near the end of the first paragraph, he reports losing hope "when I come to explain the principles, that unite our successive perceptions in our thought or consciousness." (App 20; SBN 635) It would be odd if Hume had a different thing in mind in each of these sentences. So, he seems to be acknowledging some kind of explanatory problem in the section "Of Personal Identity". But, as Barry Stroud pointed out, Hume's statement that the problem involves...
what unites or connects perceptions is ambiguous. Hume may be referring to what actually connects perceptions so that they constitute a personal bundle. But, he might also be referring to what makes us think of a personal bundle as something that persists as a unity over time, as a persisting self. Each of these interpretations is different and each suggests a very different approach to understanding what Hume found "very defective".

Stroud interprets the passage in the first way and argues that Hume's problem is that he "cannot explain how or why the 'data' from which the idea of personal identity is constructed present themselves in the way they do." Although he is surely correct in thinking that Hume cannot provide this explanation, I do not think this would have ever struck Hume as a problem. If Hume can rely on the observed fact that the imagination can progress only between perceptions within a discreet bundle, it is unclear why being unable to explain this would have worried him. Consider Hume's account of necessary connection. It too relies on an inexplicable fact: certain perceptions are constantly conjoined. These constant conjunctions produce a "determination of the mind to pass from one object to its usual attendant," (T 1.3.14.20; SBN 165) which ultimately generates the idea of a necessary connection. But why are perceptions constantly conjoined? Since every perception is a distinct existence, why does any one type stand in a constant conjunctive relation to any other? Hume's account of necessary connection requires that perceptions be arranged in these ways; it requires, for instance, that they not occur randomly. But Hume is not bothered by questions about why perceptions are constantly conjoined. Stroud's way of interpreting the

12 Stroud, 1978, pg. 133.
13 I borrow the expression "personal bundle" from Stroud (1978). It means simply bundles of perceptions roughly corresponding to what we think of as an individual person.
14 Stroud, pg. 137. See also Don Garrett, 2011. "Rethinking Hume's Second Thoughts about Personal Identity" in The Possibility of Philosophical Understanding: Essays for Barry Stroud; Cognition and Commitment in Hume's Philosophy (New York: Oxford University Press 1997), and (1981) and Peter Kail. Projection and Realism in Hume's Philosophy (Oxford: Oxford University Press 2007). I will return to these kinds of views in section V.
15 A similar complaint with Stroud is leveled by Roth (2000).
Appendix: would not be a problem that Hume could reasonably believe was unique to his account of a persisting self.\textsuperscript{16} For Hume, explanations always stop at the boundaries of experience and answering a question about why our experiences are organized the way they are, including why perceptions come in only discrete personal bundles, means stepping outside the scope of human experience.\textsuperscript{17}

Even if this is right, however, it is never something Hume explicitly says. All he does say is:

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. (App 20; SBN 635-636)

As I read this passage, Hume is claiming that he has trouble understanding why we attribute identity to a bundle of perceptions, why we take our "successive perceptions" to be a single entity persisting over time. It is important to notice that even if Hume could explain why perceptions are found only in discreet personal bundles (even if he could solve the problem raised by Stroud’s interpretation), he would still face the question of why we attribute identity to the bundle. Understanding why the imagination is confined to a subset of perceptions is not sufficient for explaining why anyone believes the self persists.

Unfortunately for Hume the simplest ways to make sense of this belief goes against principles that he cannot bring himself to reject. As he writes:

\textsuperscript{16} Garrett (1981) lodges a similar complaint but, I think, misidentifies why Stroud's problem would not bother Hume. Garrett claims that "Hume would not be dissatisfied by his inability to offer ultimate explanations for these brute causal facts; indeed, it is an important part of his philosophy that causal explanations must always lead eventually to some brute facts." (pg. 346) But, I think the reason Hume would not be dissatisfied is methodological. Hume would not be embarrassed by Stroud's concern because the impossibility of explaining why certain 'data' are present in the "universe of perceptions" as they in fact are is outside the scope of any proper philosophical explanation. We should not be surprised that explanations cannot extend to where Stroud suggests because they would then not be empirical explanations and those are the only ones Hume is after.

\textsuperscript{17} cf. Ainslie (2008)"'That perceptions come in a bundle, that introspection does not yield awareness of all the perceptions in the universe, is a brute fact. Hume shows no inclination to shy from brute facts." (pg. 151)
In short 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 never perceives any real connexion among distinct existences. Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them, there would be no difficulty in the case.\(^{(\text{App 21; SBN 636})}\)

Hume's acceptance of these principles means that his principles of association are the only way for him to explain our belief in a persisting self.\(^{(\text{19})}\) But, in the Appendix, he is beginning to see they do not. His hope vanishes "when I come to explain the principles, that unite our successive perceptions in our thought or consciousness" because he cannot think of any additional way to explain it. But why? Why can't Hume explain our attribution of identity to the self?

\textit{III}

I believe the deficiency Hume discovers becomes clear only after we understand how the principles of association work in accord with the function of a common fiction to explain its existence. As we saw earlier with the oak tree example, Hume thinks our minds wrongly take two non-identical perceptions to be a single persisting thing because of the way in which those perceptions are related. If perceptions resemble each or are causally related, the imagination moves smoothly and easily between them. And, this process feels just like that of considering a single thing over time. As Hume puts it:

That action of the imagination, by which we consider the uninterrupted and invariable object, and that by which we reflect on the succession of related objects, are almost the same to the feeling. (T 1.4.6.6; SBN 253-254)

\(^{(\text{18})}\) A classic problem with this passage is that the "two principles" are not inconsistent. It is widely agreed that Hume must therefore think there is an inconsistency between these "two principles" and something else.\(^{(\text{19})}\) See also Garrett (2011): "A standoff now seems to arise because, while still convinced that the first two alternatives cannot be right, he has suddenly lost confidence in the third and only remaining alternative."
Since the two processes “are the same to the feeling” our mind naturally confuses them and believes there is a persisting material object in the external world, e.g., an oak tree. We mistakenly think there is a single thing enduring through the changes we perceive.

The basic structure of this explanation has four steps:

Step 1: Two (or more) perceptions, A and B, stand in an associative relation R.

Step 2: Relation R in Step 1 causes the imagination to move smoothly and easily from A to B.

Step 3: The psychological process in Step 2 feels just like a different psychological process: considering an invariable and uninterrupted thing over time.

Step 4: The feeling of similarity in Step 3 generates a common fiction F.

Each step contributes to our understanding of why the mind generates F. Relation R in Step 1 causally explains the propensity of the imagination in Step 2. This transition from Step 1 to Step 2 is a fundamental aspect of human psychology (T. 1.1.6.6; SB 13). Because this feels just like considering one invariable and uninterrupted thing, the process in Step 2 causes the feeling of similarity in Step 3. But why does the feeling in Step 3 cause F in Step 4? Why does the fact that two distinct psychological processes feel the same cause our minds to produce a common fiction?

Hume cannot think that the feeling in Step 3 is causally sufficient for F because, at least sometimes, this same feeling occurs when we do consider something invariable and uninterrupted. But, in those cases we do not invent or feign fictions. Thus, in order to explain why our mind generates F, Hume appeals to the fact that F functions to resolve a type of psychological conflict. Even though the process in Step 2 feels just like the process of considering something invariable and uninterrupted, we can plainly see that A and B are not
identical. Because of this, our minds are subject to a sort of tension between what we understand, that A and B are different, and what we feel, that they are not different.

Here is how Hume describes the conflict:

The smooth passage of the imagination along the ideas of the resembling perceptions makes us ascribe to them a perfect identity. The interrupted manner of their appearance makes us consider them as so many resembling, but still distinct beings, which appear after certain intervals. The perplexity arising from this contradiction produces a propension to unite these broken appearances by the fiction of a continu'd existence. (T 1.4.2.36; SBN 205)

Or again:

Now there being here an opposition betwixt the notion of the identity of resembling perceptions, and the interruption of their appearance, the mind must be uneasy in that situation, and will naturally seek relief from the uneasiness. (T 1.4.2.37; SBN 206)

In order for the feeling in Step 3 to make sense to us and be something that we can assent to, our mind must relieve the psychological conflict. The way it does this is by feigning the notion of something persisting despite the changing interrupted perceptions, by creating a relevant common fiction. In fact, at one point Hume says that the mind forms F in Step 4 so that it can "justify" assenting to the way the psychological process feels over the understanding that it occurs between distinct perceptions.\(^{20}\) F functions to resolve the psychological conflict by allowing us to think of A and B as appearances of or aspects of a single entity.\(^{21}\) It is because it performs this important function that our minds feign F.

We can apply this structure to explain why we believe in a persisting oak tree. Two distinct perceptions of an oak tree, Oak-1 and Oak-2, resemble each other (Step 1) and this cause the imagination to move in a smooth and easy manner from Oak-1 to Oak-2 (Step

---

\(^{20}\) "In order to justify to ourselves this absurdity, we often feign some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation." (T 1.4.6.6; SBN 254)

\(^{21}\) A similar thing happens when our mind considers all of the qualities of an object at a single time. When considered gradually, we "ascribe an identity to the changeable succession of connected qualities." (T 1.4.3.3; SBN220) But when we "alter our method" of considering them, "we are presented with the idea of diversity." (T 1.4.3.4; SBN 220) Hume thinks that our mind feigns the fiction of a substance "in order to reconcile" these contradictions.
2). This feels just like considering one uninterrupted and invariable perception (Step 3).

But, our mind is not confronted with one thing; it is confronted with two things, something that it cannot ignore. Despite the fact that the passage from Oak-1 to Oak-2 feels just like considering Oak-1 for an extended period of time, it is also obviously different. This generates tension. So, the mind creates a fiction in order to alleviate this tension. It comes up with the thought of something remaining the same despite the differences in the perceptions. The belief in a Persisting-Oak, as a continued and distinct existence, accounts for the feeling in Step 3 in the face of the apparent differences between the two perceptions.

In Hume's estimation, our minds do two things requiring explanation. First, we take the psychological movement along distinct and interrupted oak-perceptions to be the same process as considering an unchanging and invariable oak-perception. We mistake the progression from Oak-1 to Oak-2 for contemplating Oak-1 over time. But, if we only did this, there would be no need for common fictions. The reason that we create the common fiction of a Persisting-Oak is because we can also understand that Oak-1 and Oak-2 are not the same, that they are not identical. So we cannot simply assent to the way a psychological progression feels because we can also see that the perceptions involved are distinct existences. The fiction of a Persisting-Oak allows our mind to reconcile the feeling of the psychological progression with the recognized distinction between Oak-1 and Oak-2.

Although I use the relation of resemblance in this example, nothing hinges on the relation being resemblance. A causal relation would work as well. Either relation would produce a smooth and uninterrupted progression of the imagination that felt like considering a single invariable thing.

Indeed, he thinks of the explanation of a common fiction as a distinct explanatory step. In the second section of Book IV, while prefacing his account of why we believe in distinct material objects, Hume tells us that he must account for four distinct things: (i) the principle of identity; (ii) a reason why "the resemblance of our broken and interrupted perceptions induces us to attribute identity to them"; (iii) an account of the propensity "which this illusion gives, to unite these broken appearances by a continu'd existence"; and (iv) an explanation for the force and vivacity of the last idea. (T.1.4.2.25; SB 200) The key thing to notice is the difference between the second and third points. In the second, Hume says he will need to explain why associative relations between "broken" perceptions induce the imagination to treat them as if they were a single entity. But in addition Hume thinks he must explain why we also believe the series of distinct perceptions to be a fictional "continu'd existence", why we believe there is something that persists through time even when we are not observing it.
When Hume attempts to apply this structure of functional explanation to the common fiction of a persisting self he encounters a serious dilemma. On the first horn, it seems that the fiction of a persisting self has no unique functional role to play. Any feeling of similarity generated by a smooth progression along a series of secondary ideas would be, or at least should be, accounted for by common fictions of material objects in the external world. Fictions of this sort relieve psychological conflicts when we have a feeling of similarity in the face of changing impressions, like Oak-1 and Oak-2. But secondary ideas differ from impressions or primary ideas only in terms of their force and vivacity, not in terms of any intrinsic qualities of the represented objects. For Hume, secondary ideas are weaker and less vivacious than the impressions from which they are ultimately derived. But if that is true we do not need a distinct common fiction of a persisting self to reconcile the way a progression of the imagination along a series of suitably related secondary ideas feels.

Suppose I remember seeing an oak tree yesterday morning (memory of Oak-1) and then remember seeing the oak tree yesterday afternoon (memory of Oak-2). The relations between these memories might cause a progression of my imagination from the memory of Oak-1 to the memory of Oak-2 (Step 2). This would produce a feeling of similarity (Step 3), a feeling as if I were contemplating one invariable and uninterrupted memory. This feeling

---

24 "Ideas produce the images of themselves in new ideas; but as the first ideas are supposed to be derived from impressions, it still remains true, that all our simple ideas proceed either mediately or immediately from their correspondent impressions." (T 1.1.1.11; SBN 7)

25 Roth (2000) describes Hume’s problems as one where "the two psychological mechanisms which respectively generate the ideas of object and of personal identity are mutually incompatible." (pg. 91) On his view, this is because "the tendencies for personal and object identity ascriptions are acting upon the same successions." (pg. 102) Once our mind runs across a series of perceptions in such a way to constitute the idea of an enduring object, it cannot, Roth argues, run across them again to generate the idea of a persisting self. I think Roth is on the right track but he is wrong to think that the psychological processes in personal and object identity "are acting upon the same successions." Hume plainly does distinguish between the token successions involved with each of these. The succession that generates our belief in personal identity contains secondary ideas and is therefore not, as Roth argues, the "same succession" as the one that makes us think there are persisting material objects. What Roth calls "psychological mechanisms" do act on successions with the same intrinsic qualities and these successions are therefore qualitatively indistinguishable; but they do not involve the same token successions.
is in clear tension with the recognizable difference between the two memories. So, my mind feigns a common fiction of a persisting self (Step 4) in order to account for this feeling of similarity in spite of the noticeable difference. The problem, however, is that Hume has already shown us that the most natural way to make sense of this feeling is by creating a fiction of a Persisting-Oak tree. When I see a tree in the morning and see it again later in the day, my mind generates the fiction of a Persisting-Oak to make sense of the feeling of similarity. But the only difference between the impressions Oak-1 and Oak-2 and memories of seeing Oak-1 and Oak-2 is the latter have less force and vivacity. This means that whatever fiction can make sense of the progression between visual experiences of an oak tree should also help with any progression between memories of those experiences. Hume has already explained the existence of common fictions whose function is to make sense of the way these kinds of psychological processes feel. The fiction of a persisting self therefore has no unique functional role.26

One might argue on Hume’s behalf that the common fiction of a persisting self does have a unique function because it is caused by a psychological progression along a series of perceptions that includes more than just memories of past experiences.27 Hume is quite clear that the bundle of perceptions constituting the self contains, in addition to memories of past experiences, pains, pleasures and other emotions (T 1.4.6.3, SBN 252). Thus, in addition to passing smoothly from a memory of Oak-1 to a memory of Oak-2, the imagination will also progress easily from, for example, the memory of Oak-1 to the memory of pleasure associated with relaxing under the tree (Step 2). If this transition feels like contemplating a

26 Again, substituting a causal relation between the perceptions would not change anything. The point is that for Hume whatever associative relations lead to the belief in a persisting self would already hold between perceptions that are not memories.
27 This is the kind of distinct explanation that Brueckner (1986) suggests Hume cannot provide. I agree but for different reasons.
single thing, it could also generate psychological tension (Step 3). However, this could not
be relieved by the mind feigning a persisting oak tree; nor could it be relieved by the mind
feigning a persisting pain. Therefore, a distinct common fiction would be needed and,
especially because the imaginative progression occurs between perceptions of perceptions
(between secondary ideas), the common fiction of a persisting self, which just is a bundle of
perceptions, appears to be especially well suited for the task (Step 4). It seems that Hume
could avoid the initial problem I raised by appealing to a distinct imaginative process that
occurs only when a person reflects on a series of her own perceptions.

This brings us to the second horn of the dilemma, which is that Hume cannot
account for a smooth imaginative transition that occurs only in cases where a person reflects
on her perceptions. This is, I think, why he says his hope vanishes when he comes to
explain the principles "that unite our successive perceptions in our thought and
consciousness." (App 20; SB636) In Hume’s view, primary perceptions are always distinct
existences and “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.” (T.14.6.16,
SBN 259) For example, we do not fictionally unite Oak-impressions with any impressions of
pleasure or of pain “in our thought or consciousness”. These are clearly distinct primary
perceptions and Hume believes they never “lose their characters of distinction and
difference.” Nevertheless, Hume also claims, “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.” (T.App.20; SBN 635) So, although the imagination does not
transition smoothly along a series of primary perceptions, an imaginative progression along a
series of secondary ideas of those primary perceptions is “felt to be connected.” (Step 3)\textsuperscript{28} Crucially, however, Hume has absolutely no resources to explain why this is true; he cannot explain why such a transition would occur only along a series of secondary ideas and not along a series of primary perceptions from which they are derived. This is once again because secondary ideas differ only in terms of force and vivacity. If the relation between my idea of pain and my idea of seeing an oak tree generates a psychological tension requiring the common fiction of a persisting self, the prior relation between my pain and my seeing the oak tree should have already required such a fiction. These latter two perceptions are more forceful and vivacious than the former and so they should stand in the same associative relations of resemblance and causality and trigger the same imaginative processes. But they do not.

The dilemma Hume faces when he tries to apply the structure of functional explanation to the common fiction of a persisting self should hopefully be clear. Either our primary perceptions are related in a way that is sufficient to cause the imagination to move smoothly between them (Step 2) or they are not. If they are, then, as we saw in the case of the oak tree, any feeling of similarity (Step 3) should be accounted for by common fictions of material objects and the belief in a persisting self is explanatorily superfluous.\textsuperscript{29} But, if they are not, there is no way for Hume to explain why the imagination would progress along a series of secondary ideas given that the primary perceptions from which they are derived are not related in ways that causes smooth imaginative transitions. Either way, there is no way for Hume to offer a complete explanation for the common fiction of the self.

\textsuperscript{28} Cf. T. 1.4.6.16
\textsuperscript{29} Hume’s difficulties here could be avoided if he had a more nuanced conception of perceptions that included a representation of a point of view. If thoughts included not simply representations of objects in the world but also a de se representation of a subject of thought, there may be a kind of psychological process that the fiction of a persisting self would be uniquely suited to make sense of. This is, however, not a conception of the nature of thought that Hume would endorse. His resistance to it, however, seems to result in serious problems. This is Strawson’s diagnosis of Hume’s shortcomings in \textit{The Bounds of Sense} (1969).
IV

In order to contrast my interpretation with others, I shall follow Garrett (2011) and Ellis (2006) in grouping existing interpretations into types. The vast number of interpretations now prohibits a comparison with each one individually within the scope of a single essay. On my view, the primary problem facing Hume's positive account of personal identity is that the conditions required for explaining our fictional ascription of identity to the self do not in fact hold. My interpretation therefore fits into the category of what Garrett calls "psychology-of-ascription" interpretations; it takes Hume's difficulty to be one of explaining why we fictionally think identity applies to the self, with explaining why we think the self persists as a single entity through time. By contrast, Garrett, like Stroud, endorses what he calls a "metaphysics-of-bundling" interpretation. Interpretations in this category understand Hume's problem to involve "how perceptions are actually linked together to make up a mind in which or on which associative mechanisms can operate." (2011, pg. 14) Most existing secondary literature falls under one of these two headings and those that do not may be collected under a miscellaneous category. Garrett further distinguishes two subcategories within the "psychological of ascription" group; first, those that think Hume's problem concerns "the operations of psychological principles other than the associative principles of resemblance and causation" and, second, those that think the problem concerns "something about the scope or operation of those two associative principles." (2011, pg. 14) The reading I am proposing is of the second kind.

This sort of division of interpretations is originally found in Jonathan Ellis. 2006 "The Contents of Hume's Appendix and the Source of His Despair," Hume Studies 5.1: 195-232. I slightly prefer Garrett's terminology. But in Ellis's terminology, "metaphysics-of-bundling" interpretations fall into 'Group 1' and "psychology-of-ascription" into 'Groups 2 and 3'; his 'Group 4' contains all the miscellaneous views.
I think this second kind of "psychology-of-ascriptioin" approach is the most promising way to interpret Hume's Appendix. Views that fall under either the miscellaneous category or the first "psychology-of-ascriptioin" subcategory cannot make very good sense of what Hume actually writes. In the Appendix he says his hope vanishes when he tries to "explain the principles that unite our successive perceptions in our thought or consciousness" and, a bit earlier in the same passage, that his difficulty involves the attempt to "explain the principle of connexion" which binds a person's perceptions together and "makes us attribute to them a real simplicity and identity." (App 20; SBN 635) It would be extremely odd if the actual problem Hume discovered had nothing to do with the principles connecting our perceptions. Since it would be best if we first tried to take Hume at his word, we should set aside interpretations that fall in either of these groups.

This, however, does not rule out the "metaphysics-of-bundling" group. In section two, I argued that Stroud's reading does not isolate a problem that would have really bothered Hume. But Garrett's own "metaphysics-of-bundling" view is slightly more sophisticated that Stroud's. Garrett argues that Hume believes the following three theses:

1) Placeless Perceptions: No non-visual and non-tactile perception is in any 'place,' either spiritual (such as a soul or mental substance) or spatial, by which it is located relative to any other perception. (2011, pg. 24)

2) Conjunctive Causation: Taken together, the following are individually necessary and jointly sufficient for the existence of a causal relation between two objects: (i) priority in

---


32 In addition to Garrett and Stroud, Kail (2007) fits into this category.
time; (ii) contiguity in time and, *where applicable*, in place; and (iii) constant similar conjunction of like objects. (2011, pg. 27)

3) Causal Bundling: Perceptions are in the same mind if and only if they are elements in a system of relevant causal relations holding among them. (2011, pg. 29)

The conjunction of these three doctrines means that any two simultaneously occurring qualitatively indistinguishable perceptions $A$ and $A'$ will always stand in the same causal relations. Since we cannot distinguish them spatially (Placeless Perceptions), it seems that $A$ will be part of a mind if and only if $A'$ is. This directly describes unacceptable metaphysical consequences. For example, Garrett notices that $A$ and $A'$ “cannot exist in different minds” and that they also cannot “be causally related to different brains and bodies.” (2011, pg. 36) This effectively eliminates the numerical distinction between $A$ and $A'$. If $A$ and $A'$ cannot be qualitatively or spatially distinguished, by the indiscernability of identicals, $A$ and $A'$ are identical. No two similar and simultaneous perceptions could belong to different minds.

A simple way for Hume to avoid Garrett’s problem would be to reject the sufficiency of Conjunctive Causation. He could maintain that priority and contiguity in time as well as constant conjunction are necessary for perceptions to stand in a causal relation but deny that they are sufficient. Thus, $A$ and $A'$ may exist simultaneously without standing in the same causal relations, which means they need not belong to the same personal bundles. But this response to Garrett’s problem is not ideal because it is inconsistent with Hume’s first

---

Garrett (2011), pg. 33. This is sort of problem is presented with less detail in his (1981) and (1993). In his (2011) Garrett raises two additional unacceptable consequences that have to do with the whether perceptions could be causally integrated into discreet bundles. But, as with Stroud's interpretation, it seems to me that Hume’s observation that causally discreet personal bundles of perceptions exist is sufficient to avoid these consequences. Since there are distinct personal bundles of perceptions with causal relations holding between the members of a single bundle, Hume would be troubled by these sorts of worries about the causal integration of perceptions.
It is, however, consistent with his second definition wherein two things are causally related only if "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; SBN 170) Nevertheless, this way of dodging Garrett’s concerns is not ideal.

It seems clear that Hume would face a serious problem if he were committed to denying the obvious fact that different minds often have the same ideas. But it is not clear to me that he is. Garrett is right to say that Hume’s view entails that no two qualitatively indistinguishable and simultaneous perceptions can belong to different minds. However, contrary to what Garrett indicates, the unacceptable conclusion that two minds could not have the same idea does not follow from this. It is just that they must literally share the same idea, instead of two ideas of the same type. According to the bundle theory, two minds would share a perception just in case the bundles that constitute those minds partially overlap. For instance, when you and I have the same idea of an oak tree it is because a single idea belongs to both your bundle and to mine. This seems to be a straightforward consequence of Hume’s bundle theory and, although it is certainly odd, I do not believe it generates worries that would have bothered Hume in the Appendix. The metaphysical consequences Garrett highlights might give us reasons to resist accepting the bundle theory of mind, but, for someone like Hume who endorses the theory, they do not seem to present any additional problems that Hume might have failed to notice. It seems to me that Hume could simply accept Garrett’s point that no two qualitatively similar and simultaneous perceptions could belong to different minds. It is far less counterintuitive within Hume’s own system than Garrett makes it out to be.

34 "We may define a cause to be 'an object precedent or contiguous to another, and where all the objects resembling the former are plac’d in like relations of precedency and contiguity to those that resemble the later" (THN 1.3.14.31 SBN 170)
This leaves us with the second "psychology of ascription" group. Garrett objects to interpretations in this category by arguing that they do not locate a problem that "Hume would have thought to pose a serious problem or 'contradiction' had he seen it." Even if Garrett is right in thinking that this is a shortcoming for some if the existing interpretations, it is not a difficulty for the one I am advancing. The dilemma I raised for Hume stems from the simple fact that either primary perceptions are causally related or they are not. If they are, then the feeling of similarity generated by the smooth progression of the imagination can be explained by common fictions of persisting material objects (as in the oak tree example). But if they are not, then Hume cannot explain why any secondary ideas of these perceptions are causally related. On this second horn especially, it would make sense that Hume might have worried about the possibility of a "contradiction". The generality of Hume’s principles of association means that the imagination should transition between any two suitably related perceptions. Thus, two secondary ideas X’ and Y’ are related in such a way as to cause the imagination to smoothly transition from X’ to Y’ if and only if primary perceptions, X and Y, (from which they are derived) are related in such a way as to cause the imagination to smoothly transition from X to Y. However, in the case of the self, X and Y are not so related; a contradiction threatens because X’ and Y’ are. If there is any actual logical “contradiction” that Hume could be worried about, surely this is it. Hume could

---

35 (2011, pg. 15). This is the same objection that Ellis (2006) raises to psychology of ascription approaches that are not his own. Ellis’s own interpretation is intriguing but it misrepresents Hume. Ellis suggests a subtle problem with Hume’s account of the idea of the self stemming from the fact that there is no impression from which the idea can be copied. But Hume is not trying to offer an account of the idea of the self; he is trying to explain why we believe the self persists or why we attribute identity to it. Hume acknowledges the "true idea" of the self prior to his explanation of a fictional ascription of identity to it. Cf. Garrett (2011).

36 It is, I think, a correct criticism of Haugeland (1998) and Winkler (2000). Other recent interpretations in this group include Ainslie (2001) and (2008), Baxter (2009), Ellis (2006) and Roth (2000). Although I favor the interpretation presented here, Ainslie (2008), Baxter (2009), and Roth (2000) have comparatively minor problems.
avoid the contradiction but then he would face the problem posed by the first horn of the dilemma.

But is this dilemma the problem Hume is referring to in the Appendix? What about other interpretations in the "psychology-of-ascription" category? As I mentioned earlier, I find these to be the most plausible but it seems that the vagueness of the Appendix prohibits determining whether any one succeeds in capturing what precisely Hume had in mind. However, my interpretation has one significant advantage over others. It describes a difficulty that is closely connected to a fundamental tenet of Hume's entire philosophical system. There is a very important symmetry between Hume's account of personal identity and his accounts of other common fictions. Other interpretations of the Appendix neglect this, isolating Hume's discussion of personal identity. This is regrettable because since the publication of the Treatise, even though Hume's views on specific issues have been quite thoroughly discussed, his overall philosophical project receives comparatively little discussion from commentators. In this paper, I have attempted to make the connection between Hume's explanatory functionalism and his thoughts on personal identity more perspicuous because I do not think we cannot adequately understand the his overall project in the Treatise until we better understand the features his various explanations share.

---

37 To be fair, some do try to draw connections between Hume's account of personal identity and other aspects of his philosophy. Ainslie (2001) argues that the account is connected to other central ideas in Book I Part 4 and Baxter (2009) thinks that the problem Hume faces is with his more general account of identity.

38 Acknowledgements.