

THE SYNTHESIS OF CONCEPTS: INFERENTIALISM AND SEMANTIC THEORY
IN HUME, KANT, AND HEGEL

David Landy

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Philosophy.

Chapel Hill
2008

Approved by:

Alan Nelson

Jesse Prinz

William Lycan

Gerald Postema

Marc Lange

ABSTRACT

DAVID LANDY: *The Synthesis of Concepts: Inferentialism and Semantic Theory in Hume, Kant, and Hegel*
(Under the direction of Jay Rosenberg and Alan Nelson)

I re-cast the history of Modern philosophy as a debate about the nature and content of mental representations, a debate that is first made explicit by Hume, and which crescendos with the contrasting theories of Hume and Kant. Hume is a sophisticated relationalist who believes that content is fixed by a relation between a mental entity and that which it represents. Kant, on the other hand, rejects relationalism on the grounds that it makes impossible our representing as such a world of objects bearing lawful relations to one another. Since he argues that this is necessary for representing oneself as a single, unified subject of experience persisting through time, he concludes that relationalism is untenable. Kant presents inferentialism—the thesis that the content of a representation is constituted by that representation’s role in a system of inference—as a viable alternative to relationalism. Hegel accepts the Kantian picture, emphasizes the normativity involved in the inferential articulation of concepts, and argues that this is an essentially social affair. By reading these figures in this way I am able to reveal the motivations behind their semantic programs and uncover arguments that have been underappreciated in scholarship on Modern philosophy and in contemporary semantic theory. The most significant of these is Kant’s argument from the necessary co-representation of self and world to an inferentialist theory of conceptual content.

ACKNOWLEDGEMENTS

I would like to thank all of those who acted as my advisor during the writing of this dissertation. First and foremost was Jay Rosenberg, who was an outstanding advisor as well as my model for all things philosophical. Next is Jesse Prinz who acted as my much-needed antagonist and advisor pro tem in Jay's absence. Finally, Alan Nelson who oversaw the final stages of the dissertation and helped and encouraged throughout. I am indebted, also, to Bill Lycan, Marc Lange, Jerry Postema, and Don Garrett for their invaluable guidance and feedback.

I would also like to thank the philosophical community at the University of North Carolina at Chapel Hill for the support and counsel that they have given me; I wish especially to thank Margo Chiovoloni, Drew Johnson, Piers Norris Turner, and Elizabeth Foreman for their patient comments on and discussions of numerous versions of this work.

I also wish to thank my mother, Judy, and my father, Rick. Without their love and support this dissertation would not have been possible.

TABLE OF CONTENTS

Chapter	Page
I. HUME'S IMPRESSION-IDEA DISTINCTION	1
II. A (SELLARSIAN) KANTIAN CRITIQUE OF HUME'S THEORY OF CONCEPTS	34
III. INFERENTIALISM AND THE TRANSCENDENTAL DEDUCTION	52
IV. SELLARS ON HUME AND KANT ON REPRESENTING COMPLEXES.....	85
V. HEGEL'S ACCOUNT OF RULE-FOLLOWING.....	117
VI. A KANTIAN ACCOUNT OF RATIONAL THEORY- CHANGE.....	150
BIBLIOGRAPHY.....	179

Chapter One

Hume's Impression-Idea Distinction

Hume sets out in the first book of the *Treatise* to present a theory of the mental according to which everything mental can be accounted for in terms of mental entities and their relations and behaviors. Hume calls such mental entities "perceptions" and divides these into two important classes: impressions and ideas. Hume's official position on what determines whether a mental entity is an impression or an idea has been the subject of some debate, largely because Hume's text is subtly ambiguous on the issue. One main point dividing scholars has been how to treat the degree of force and vivacity that impressions and ideas have—whether to take this as the determinative criterion for what makes impressions and ideas what they are, or as a mere symptom, helpful for distinguishing the two in introspection, but not constitutive of what it is to *be* either an impression or an idea.¹ I will argue that force and vivacity are best understood as phenomenal *symptoms* by which we recognize a distinction that is best *explained* by

¹Stroud, (1977) and Noonan, (1999) are both proponents of the first option: the force and vivacity interpretation. Bennett, (1971) reads Hume as employing force and vivacity in this way, but argues that to make Hume's theory plausible we must switch to using an objective realm interpretation. Everson, (1988), recognizing the difficulty with all of these approaches reinterprets force and vivacity functionally, and argues that this reading avoids many of the usual problems with the force and vivacity reading. Garrett, (1997) argues that the force and vivacity interpretation must be supplemented with one that makes crucial use of the Copy Principle, and that both criteria are important for understanding Hume's distinction. Dicker, (1998) argues that Hume is himself ambiguous on the point, and that neither kind of reading is a particularly fruitful in any case.

Hume's Copy Principle, so that the strictly speaking, the distinction between ideas and impressions is drawn using that principle, and *not* force and vivacity.²

I will argue, that is, that Hume observes that we seem to be very good at—although we make mistakes at times—sorting our perceptions into two classes—impressions and ideas—and that what *explains* our ability to do so is the fact that there really *is* a distinction between these two *kinds* of things: one is a copy of some other mental entity (ideas) the other is not (impressions). I will further argue that, for Hume, what it is to be a copy is to meet two necessary and *jointly* sufficient conditions. The first is that a copy is always caused by—in the sense that Hume makes of this notion after the proper investigation undertaken later in the *Treatise*—that of which it is a copy. The second is that a copy always exactly resembles that of which it is a copy. Failure of an entity to meet either one of these criteria means that that entity is *not* a copy. (It is worth noting at the outset that Hume's is a rudimentary analysis of "copy", whose merits and faults could be the topic of an investigation distinct from the current one; we will proceed using Hume's notion.)³ Impressions always so fail in one way or the other and so are not copies. Ideas always meet *both* criteria and so are copies. This is the real distinction upon which we touch when we intuitively sort our perceptions into these two classes. I will

²This is not meant to imply that the force and vivacity reading and the Copy Principle reading are the only two interpretations available. Rather than discuss all the alternatives, I will focus on showing the disadvantages of the former and the advantages of the latter. One such alternative that is worth mentioning is that which draws the impression/idea distinction by relying on an intuitive feeling/thinking distinction. Suffice it to say that my position is that insofar as the feeling/thinking distinction is an intuitive one, it is involved in the pre-theoretical sorting of mental entities which provides the explanandum of which the Copy Principle is the explanans.

³Ruth Garrett Millikan gives a much more robust and accurate account of copying in Millikan, (1984). It would be an interesting topic of investigation to see how Hume's account of the impression/idea distinction would fare using that account of copying.

begin my investigation of this issue, however, by examining the force and vivacity interpretation.

The greatest advantage of reading the distinction between impressions and ideas as being constituted by the degree of force and vivacity that mental entities have is that this seems to many to be the most straightforward reading of Hume's actual text. Hume writes, for instance, that,

All the perceptions of the mind are of two kinds, *viz.* impressions and ideas, which differ from each other only in their different degrees of force and vivacity. (T 1.3.7.5; SBN 96)

If the *only* difference between impressions and ideas is their degree of force and vivacity, the only candidate available for that which makes a mental entity an impression or an idea is degree of force and vivacity. I will return to this particular passage later to explain how it can be read so as to be consistent with drawing the distinction between impressions and ideas using criteria other than degree of force and vivacity. For present purposes, however, it will be sufficient to point out that whatever advantages such passages lend to the interpretation we are considering are at least counterbalanced (if not outweighed) by other passages in which Hume relies on some *other* criteria for drawing this distinction. For instance, in the *Enquiry* Hume writes of ideas that,

except the mind be disordered by disease or madness, they never can arrive at such a pitch of vivacity, as to render these perceptions altogether undistinguishable. (EHU 2.1; SBN 17)

This implies that in the mind that *is* diseased or mad ideas *can* obtain a degree of force and vivacity equal to that of impressions. For it to be possible, however, to say that some idea has the same degree of force and vivacity as some impression, there must be some criterion *other than* force and vivacity that makes these different kinds of mental entities.

Thus far, Hume's texts on these issues seem to be ambiguous, and so I now want to turn to more *philosophically* motivated reasons for thinking that there are enough serious problems with the force and vivacity reading to motivate looking elsewhere. (We will have cause to return to some further texts a bit later.) We can proceed, for now, under the assumption that the best way to understand Hume's distinction is whatever way, in accordance with the text, makes his philosophical position strongest. As the last passage demonstrated, one encounters the most serious difficulties with the force and vivacity interpretation in cases where what we, and Hume, would intuitively think of a particular mental entity as an idea, but where its degree of force and vivacity is clearly equal to, or greater than, that of some corresponding impression.⁴ Barry Stroud provides a particularly compelling example of just this sort. Stroud imagines a detective who upon first examining a murder-scene finds nothing out of the ordinary. Remembering the scene later, the detective recalls that there was a poker leaning on the left-hand side of the fireplace, despite the fact that the victim was right-handed. This fact suddenly stands out to the detective as the key to solving the crime because he realizes that since so-and-so is left-handed, he must be the murderer.⁵ The moral of the story is, of course, that the *memory* (a paradigm example of an idea for Hume) of part of the scene of the crime is more vivacious to the detective than was the impression of it, but that in order to express this we must rely on some criterion for distinguishing impressions from ideas other than force and vivacity. What makes this approach even more difficult as an interpretation of Hume is that Hume himself is aware of such counterexamples, and seems to retreat from

⁴This, of course, supposes that we have some intuitive, pre-theoretical grasp on the distinction between impressions and ideas. I will argue later that this is a claim on which Hume heavily relies.

⁵ Stroud, (1977): 28.

using force and vivacity as criterial when considering them, as we saw him do in the passage from the *Enquiry* above. So it seems that Hume does not, and ought not, use force and vivacity as the criteria for determining whether a mental entity is an impression or an idea.⁶

If, then, we are to reject degree of force and vivacity as the criterion for determining whether a mental entity is an impression or an idea, to make sense of this distinction, we will need some other criterion to do this work. Again, however, even amongst those who agree on this much, scholarly opinion is divided. Bennett, for instance, is a proponent of the view that impressions of sensation, at least, *must* be explained as being those mental entities that first make their appearance as a result of people veridically perceiving physical objects.⁷ This view, however, comes with difficulties of its own, one of which is particularly pressing. This is that, as is well known, the very notions of physical objects, perceiving physical objects, and veridically perceiving physical objects all become problematic (at the least) later in the *Treatise*—in no small part *because* of Hume’s theory of impressions and ideas—and so provide an unstable and awkward foundation on which to interpret Hume as building. That is, if the notion of veridically perceiving physical objects is to ground the distinction between impressions and ideas, but that notion turns out to be bankrupt later in the *Treatise*, then we would also be forced to give up this distinction. That, however, seems far too high a price to pay. Furthermore, Hume’s official line on the origins of impressions—found in

⁶There are, of course, moves that can be made to save the force and vivacity reading in the face of such examples. The point here is just that such examples, even if surmountable, give us some reason to look for other ways of drawing this distinction.

⁷It is important to note that Bennett does not read Hume as drawing the distinction this way (he is a force and vivacity proponent *qua* Hume interpreter), but thinks that if Hume’s theory is to be at all taken seriously, we must revise his stance here for him in just this way.

close proximity to his discussions of the distinction between impressions and ideas—is that impressions of sensation come from “unknown causes” (*T* 1.1.2.1, SBN 7). Hume steadfastly refuses to speculate on the origins of impressions, and so any interpretation of the distinction between impressions and ideas that relies essentially on positing some particular origin for these cannot be one that we attribute to Hume without a great deal of reservation.

Stephen Everson recognizes this difficulty and takes it as a condition on an acceptable interpretation of Hume that the distinction between impressions and ideas be drawn entirely without reference to anything non-mental, a condition I endorse. Everson infers from this that another condition on any such account is that “one be able to distinguish impressions and ideas introspectively”⁸ and he takes it that Hume thought the same. If true, this would be reason to reconsider the force and vivacity interpretation, despite its difficulties, as force and vivacity clearly does meet this latter requirement. Everson does just this. It is important to see, however, that this second condition does not, in fact, follow from the previous one unless one *also* attributes to Hume the thesis that everything mental is available to introspection. That is, it does not follow from the fact that the distinction between impressions and ideas must be drawn without reference to anything non-mental that the distinction thus drawn must be introspectively available, *unless* it is also true that everything mental is transparent to introspection. If something’s being available to introspection means that it can be known infallibly, we must reject this further condition.⁹ Given, for instance, Hume’s *error theory* concerning our concept of

⁸Everson, (1988): 401-413, 404.

⁹If, on the other hand, being available to introspection does not imply infallibility, but only that what is mental can appear to consciousness, then this condition is, of course, fine.

causation, for example, attributing this supporting premise to Hume would clearly be a mistake. That error theory clearly demonstrates that Hume thinks that there are at least *some* mental states that are *not* infallibly available to introspection. It follows, then, that availability to introspection, in this sense, cannot be a necessary condition for a state's being a mental state. Still, we can agree with Everson, against Bennett, that since Hume wishes to remain uncommitted about the source of impressions, we cannot *define* these in a way that makes essential reference to this source.

Of course, this transparency thesis is not without its own, independent support in Hume's texts. For instance, in the section of the *Treatise*, "Of skepticism with regard to the senses," Hume discusses the thesis that our senses represent the world as distinct from the mind. He writes,

Add to this, that every impression, external and internal, passions, affections, sensations, pains and pleasures, are originally on the same footing; and that whatever other differences we may observe among them, they appear, all of them, in their true colours, as impressions or perceptions. And indeed, if we consider the matter aright, 'tis scarce possible it shou'd be otherwise, nor is it conceivable that our senses shou'd be more capable of deceiving us in the situation and relations, than in the nature of our impressions. For since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Every thing that enters the mind, being in *reality* a perception, 'tis impossible any thing shou'd to *feeling* appear different. This were to suppose, that even where we are most intimately conscious, we might be mistaken. (T 1.4.2.7; SBN 190, my emphasis)

Hume strong language here certainly seems to imply that he endorses some sort of transparency thesis, and one that implies a kind of infallibility as well. He has argued earlier that it is impossible for our senses to deceive us about the nature of our perceptions. Here he claims that it is equally impossible that they should deceive us about the situation and relations of such perceptions. He claims further that because consciousness is aware of every perception, we necessarily see such perceptions *as*

perceptions, not as something distinct from the mind (as we would have to in order, Hume claims, to obtain the idea of an external world).

What is crucial in understanding this passage is that one keep in mind that Hume is here arguing that *the senses* are not the source of our mistake about the distinction between the external world and our perceptions. He is decidedly *not* arguing for the thesis that no such mistake is possible. *That* would make Hume's puzzlement over the source of such a mistake utterly inexplicable. In fact, this passage occurs as part of an argument *for* the conclusion that it is "some inference either of the reason or imagination" that is the source of this error. Of course, what this means is that while it may be that the *senses* are not the source of any error about our mental states, there are *other* sources for such error. Thus, introspective infallibility still cannot be a necessary feature of mental processes, for it may be the case that some feature of our mental lives other than our senses deceives us about the nature, situation or relations of our perceptions.

Furthermore, in the second half of this passage, in which Hume claims that it is impossible that we should ever be deceived about what appears to consciousness, it is important to notice that what Hume is here explicitly addressing is the possibility that we should ever think that something presented to consciousness is a *non-mental entity*. Hume's claim is merely that what appears to consciousness is necessarily mental, and can be known to be such. This follows almost immediately from Hume's views on consciousness and mentality. This thesis does imply that all mental entities can be known, via introspection, to be such. It does not, however, imply anything about whether the properties or relations in which such entities stand can be so known. With that said, I take it that the best evidence that Hume held the thesis that everything mental can be

infallibly known via introspection has been defeated. Thus, we can safely endorse Everson's condition that the criterion for distinguishing impressions from ideas must make essential reference to anything non-mental, while still resisting his claim that this distinction must be available to introspection.

We now have one condition in place that our alternative criterion for grounding the impression/idea distinction must meet, and we have rejected another. At this point, it will be best to turn once again to the texts for our next clue. In particular, the following two passages concerning ideas and impressions from the *Enquiry* and the *Treatise*, respectively, will be of some help.

[A]ll our ideas or more feeble perceptions are copies of our impressions or more lively ones. (EHU 2.5; SBN 19)

[A]ll our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly resemble. (T 1.1.1.7; SBN 4)

What is noteworthy about these passages is that, while we find Hume in the first passage again gesturing at the difference in degree of force and vivacity of impressions and ideas, we find him in both passages also emphasizing the genitive difference between the two kinds of mental entities—i.e., he is pointing out that ideas are *copies* of impressions. While many readers of Hume take this to be a mere fact *about* impressions and ideas, it is this difference that I propose we take as the *criterion* that determines whether a mental entity is an impression or an idea.¹⁰ Impressions are the original objects of the mind, derived from sources unknown; they are not copies of any other mental entities. Ideas are copies, either of impressions or of other ideas. It is this difference that makes a perception either an impression or an idea.

¹⁰At least with regard to *simple* impressions and ideas.

With that said, a number of tasks have suddenly made their way on to our agenda. The first will be to explicate the key notion here of what it is, for Hume, for one mental entity to be a copy of another. The next will be to address objections to using this particular criterion for distinguishing impressions from ideas. The last will be to return to the texts that seemed to support the force and vivacity reading of this distinction to show how they are also at least compatible with the proposed criterion.

To begin, then, we can notice that in seeking evidence for his Copy Principle—the thesis that all ideas, are, in fact, copies of impressions, or copies of copies, etc.—Hume relies on two kinds of evidence: evidence showing that ideas exactly resemble their corresponding impressions, and evidence showing that ideas are caused by their corresponding impressions.¹¹ So Hume cites as evidence for this principle first that,

The first circumstance, that strikes my eye, is the great resemblance betwixt our impressions and ideas in every other particular, except their degree of force and vivacity [...] the ideas I form are exact representations of the impressions I felt; nor is there any circumstance of the one, which is not to be found in the other. (T 1.1.1.3; SBN 2)

And next that,

The constant conjunction of our resembling perceptions, is a convincing proof, that the one are the causes of the other; and this priority of the impressions is an equal proof, that our impressions are the causes of our ideas, not our ideas of our impressions. (T 1.1.1.8; SBN 4)

A word is in order on each of these conditions. Firstly, it is worth noting that the Exact Resemblance Condition applies only to the “circumstances” of ideas and impressions, where “circumstance” is best understood here as what we might call its characteristic, or imagistic, quality. The thought here is that impressions and ideas have a certain imagistic quality—be it a particular color, shade and hue, or a note, timbre and tone, etc.—without

¹¹In what follows, I will be drawing on Don Garrett’s explication of the Copy Principle in Garrett, (1997).

which that perception is not the impression or idea that it is, but rather an impression or idea *of* something else (if impressions can be said to be *of* anything). For instance, one's idea of some painting, say, is only an idea of *that* painting if it exactly resembles that painting; change a brushstroke here and a color there, and one now has an idea of some other, slightly different, painting. So the Exact Resemblance Condition is actually that a copy must exactly resemble its original in all essential qualities.¹² This importantly excludes degrees of force and vivacity, as the above quotation makes clear, and as we will discuss later, relational properties holding between mental entities.

Regarding the Causal Condition, one might object here that just as we rejected the objective-realm account of impressions because the notion of an objective realm becomes problematized for Hume, so should we be hesitant to place *causation* in so central a role in our own account because this becomes equally problematized later in the *Treatise*. The key here, though, is that the two notions are not *equally* problematized. Hume offers an *error theory* regarding our concept of causation, complete with an account of how we can justifiably put that concept to use. He offers no such alternative for 'substance', the key problematic term in the objective realm interpretation. And as the above quotation clearly demonstrates, it is only the de-problematized notion of causation that actually does come into play when Hume employs this condition. Without delving into the issues surrounding Hume on causation too deeply, we can note that Hume offers a two-part alternative construal of the concept of causation:

¹²Notice that this rules out, at least in the vast majority of cases, that a complex idea is a *copy*. A more precise formulation would instead be that a complex idea is an idea composed of simple ideas that are themselves copies of simple impressions. I think this is exactly what Hume wants to say about complex ideas. I am indebted to Peter Millican for prompting me to make this more clear.

An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in a like relation of priority and contiguity to those objects, that resemble the latter. (T 1.3.14.31; SBN 170)

An object precedent and contiguous to another, and so united with it in the imagination, 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. (T 1.3.14.31; SBN 170)

Clearly, at least the first part of this two-part construal is operative in Hume's use of the Causal Condition, and arguably the latter is as well. Hume's argument for the Copy Principle relies on exactly the premise that ideas and impressions are constantly conjoined, with impressions always preceding their correspondent ideas. Thus, the distinction between impressions and ideas is not undermined by placing causation, *properly construed*, at the center of Hume's account, which is just what Hume does. It would only be so undermined if it was the problematic notion of causation—as a metaphysically necessary connection that outruns regular succession—that was used, which it is not.

What both of these conditions reflect, which will be important for us to keep in mind throughout the course of our investigation, is that what Hume is concerned with in speaking about perceptions is the nature of the entity that does the representing, not that which is *object* of this representing. In the specific case of perceptions, this object is a certain mental entity: an impression or idea. Thus can Hume speak of the *cause* of a perception, for example. This is shorthand, in Hume, for the cause of the *existence* of the perception. Similarly, when Hume speaks of the exact resemblance of one perception to another, he is speaking of the exact resemblance of one mental *entity* to another mental *entity*. Finally, this has important consequences for how we conceive of force and vivacity. One *might* think that these are features of the object of representation, but this, I

think, is pretty clearly wrong of Hume's account. I will come back to this subject later, but this much of a telegraphic remark will be appropriate here: conceiving perceptions as mental *entities* allows us to construe force and vivacity as non-essential, non-relational qualities of such mental entities. That this ontological commitment is at the core of Hume's semantics is important to keep clear about.¹³ We will have more to say about the *object* of representation—and how such representations come to so much as *have* objects—later.

So, the Copy Principle states that all (simple) ideas exactly resemble in their intrinsic qualities, and are caused by, some corresponding impression. It is worth pausing for a moment in our dialectic to note that, qua an account of what it is for one thing to be a copy of another, Hume's Copy Principle leaves much to be desired. While Hume's two conditions might be necessary conditions on copies—although the exact resemblance condition is probably too strong as it stands—it is hard to imagine that they would also be sufficient. For instance, there are certainly cases in which one thing might be the cause of another, and *just so happen* to exactly resemble it, but nonetheless the latter would not be a copy of the former. Ruth Garrett Millikan, in *Language, Thought and Other Biological Categories*, spends a good deal of time presenting her own account of copying, which—although she disavows that this account is anything more than a definition of a homophonic term of art—is much closer to an adequate analysis than is Hume's.¹⁴ Nonetheless, Hume's definition will do for the purposes of a rudimentary account of the mental, to be supplemented as is needed by those who take up the task of keeping his research programme contemporary.

¹³Kant, for instance, later rejects this *ontological* conception of accounting for meaning.

¹⁴Millikan, (1984).

It is also worth pausing to address a potential confusion about Hume's account of copying. Hume's conditions—exact resemblance and causation—are, as Hume's conceives them, necessary and only *jointly* sufficient. If a perception, or any potential copy, fails to meet *either* of the conditions, then it is not a copy. So, for example, suppose the phone rings twice. The second ring exactly resembles the first. However, since the second ring is not *caused by* the first—something that Hume certainly wants to be able to say—the second is not a copy of the first. Suppose further that I throw a baseball through a window, and the window breaks. My throwing of the baseball is the *cause* of the window breaking, but because the window's breaking does not exactly, or at all, resemble the throwing of the ball, the former is not a copy of the latter. There are two ways, that is, for something to *fail* to be a copy. It can *either* not exactly resemble that which causes it, *or* not be caused by that which it exactly resembles. This will be important to keep in mind, especially when we reach our discussion of impressions of reflection.

Hume's version of the Copy Principle, then, states that all (simple) ideas exactly resemble in their intrinsic qualities, and are caused by, some corresponding impression. This is a familiar claim of Hume's, although, as mentioned earlier, it is not usually put to use in quite the way we are doing here. It is usually presented as a claim *about* impressions and ideas, not as a criterion for what it is *to be* an impression or an idea. Drawing the distinction between impressions and ideas this way, employing the Copy Principle to do this work, means that we must now face a cluster of issues in Hume scholarship over which there has, again, been some disagreement. We can begin with an objection to drawing the distinction this way, and will soon be off and running; the objection runs as follows. If we draw the distinction between impressions and ideas along

these lines, don't we make Hume's citing of empirical evidence in support the Copy Principle, and his claim to be willing to entertain *empirical counterexamples* to the Copy Principle utterly mysterious and/or nonsensical? That is, if "impressions" and "ideas" are *defined* as original mental entities and copied mental entities, respectively, how could empirical evidence ever come to bear one way or the other on the claim that impressions are original mental entities, etc.? On the other hand, scholars have wondered, if the Copy Principle is merely an empirical claim, and Hume is genuinely prepared to consider counterexamples to it, how can he use it to *refute* the claims of other philosophers that contradict it? On what grounds do these claims count as refuted rather than as counterexamples to the Copy Principle itself? Furthermore, on a somewhat different though clearly relevant note, what the heck is going on with the missing shade of blue, which Hume admits as an *actual* counterexample to the Copy Principle?!

The best way to begin to answer these questions is to look at the process Hume goes through in proposing, defending, and employing the Copy Principle, and to take note of just what Hume is up to at each stage. Hume begins this process by calling his readers' attention to a distinction—the distinction between impressions and ideas—of which he thinks we all already have an intuitive grasp. He writes that,

[I]t will not be very necessary to employ many words in explaining this distinction. Every one of himself will readily perceive the difference betwixt feeling and thinking. (T 1.1.1.1; SBN 2)

We can all pretty well sort our mental entities into impressions and ideas already, even if we have never used these words to describe them, or thought about doing so at all explicitly. Just to make sure of this, and to focus our attention on the right distinction Hume cites some paradigmatic examples of each (sensations, emotions and passions in

the case of impressions, memories in the case of ideas), and points to certain phenomenal qualities (degree of force and vivacity) by which each is commonly recognized.¹⁵ So far, the only important philosophical work done is to have called our attention to this distinction with which we are all already, Hume presumes, familiar.

Next, Hume moves through the simple/complex distinction, which becomes important soon after, but which we can skip for the moment. He then announces that,

Having by these divisions given an order and arrangement to our objects, we may now apply ourselves to consider with the more accuracy their qualities and relations. (T 1.1.1.3; SBN 2)

That is, having drawn our attention to and made clear what the distinctions are that are to be the object of the current inquiry we can get to the business of investigating what grounds them. That is, again, we are all capable of sorting our perceptions into impressions and ideas. Hume now wants to consider if our so sorting our perceptions corresponds to any *real* difference of kind among those perceptions (other than that some are the kind that get sorted as impressions and others are of the kind that get sorted as ideas). Hume is concerned with what, if anything, *accounts for* our sorting of perceptions into these two classes. The question that is now before Hume is what qualities must a mental entity have if it is to be an impression or an idea (if there is to be such a distinction), and how is this distinction related to the simple/complex one. This, of course, is exactly where Hume first proposes the two criteria that make up the Copy Principle; a copy must exactly resemble its original, and it must be caused by that

¹⁵It is important to be clear here. It is not the case that focusing our attention on this distinction is the *only* work that is to be done by force and vivacity in Hume's system. Clearly it is not. Force and vivacity also play a crucial role in Hume's theory of belief, his theory of judgments (insofar as he offers one), etc. It is simply that force and vivacity, in the context of drawing the distinction between impressions and ideas is not criteriological, but symptomatic.

original. So, the Copy Principle is first introduced by Hume as a way of accounting for a distinction on which we all already have some intuitive grasp.

Accordingly, Hume's next step is to provide evidence showing that the distinction with which we are all already familiar *can* be accounted for in this way, that is, that our pre-theoretical sorting of mental entities into impressions and ideas matches up with those mental entities that are original and those that are copied. As we have already seen, he does this in two stages corresponding to the two aspects of the Copy Principle. First he notes that by introspecting he has discovered that ideas do seem to exactly resemble impressions, and next he presents evidence that ideas are also caused by impressions. Hume has now presented evidence that shows that (a) ideas are all copies of impressions, and thereby, (b) this is what grounds our intuitive, pre-theoretical sorting of our mental entities into these two classes; i.e., it is this difference that *accounts for* our sorting our perceptions this way. For instance, it is *because* ideas are copies and impressions are not that the former are less forceful and vivacious than the latter, and it is that fact, in turn, that helps us distinguish one kind of perception from the other introspectively as we do. (The gist of this explanation is that when ideas are copied from impressions some, but not all, of the force and vivacity from the latter is transferred to the former. So, copying plays a crucial role not only in constituting the real distinction between impressions and ideas, but also thereby in accounting for the phenomenal quality of this distinction. This is exactly what a good account of what underlies appearances ought to do.)

The Copy Principle, according to this story, is something like a *theoretical explanatory* principle. We start out with a distinction with which we are all familiar. We then wonder if there is anything really to this distinction, if our sorting of mental entities

corresponds to any real difference between them. We discover that it does. What we intuitively place in the category of ideas are all copies of what we intuitively place in the category of impressions. So, the distinction at which we were getting all along (via differences in degrees of force and vivacity, for example) is really the distinction between copied mental entities and original ones. In fact, even our being able so to use, for example, force and vivacity to recognize this distinction is accounted for by this theoretical explanatory hypothesis—as noted above. As science—and keep in mind that Hume takes himself to be conducting the science of man—often does, Hume’s theory is one that explains not only what underlies the appearances, but also why the appearances are as they are. Of course, this can *only* be the case *if it is true that* ideas are copies of impressions. So Hume’s task is two-part; first he must show *that* ideas are copies of impressions, and second he must show that *it is because* ideas are copies of impressions that we can sort them as we do. In this sense, there is both an empirical component to the Copy Principle as well as a criterial one. Via empirical means, the Copy Principle is discovered to be true, and it is thenceforth employed in an explanatory and criterial role. The principle is *justified* empirically, and *employed* explanatorily. In those first paragraphs of the *Treatise*, then, Hume presents the evidence in support of the criterial component *by* presenting evidence for the empirical component.

Having presented that case, Hume is now honest enough to present the case *against*—the missing shade of blue. Before we move on to discussing this example in more detail, however, we must stop to explain what critics of Hume have drawn attention to: namely, that Hume does not here consider the claims of his philosophical opponents

that contradict the Copy Principle as counterexamples to his proposed way of accounting for the intuitive distinction between impressions and ideas.

What Hume has done so far is to make an empirical claim—that all ideas are copies of impressions—and a theoretical one—that this empirical claim gives us good reason to suppose that what grounds our intuitive distinction between impressions and ideas is that the latter are copies of the former. The objection now on the table is that his original empirical claim is not well-founded, that there are some ideas that are not, and could not be, copied from any impression. Instead of taking his *empirical* claim to be refuted, the critic complains, Hume seemingly rejects the very possibility of such counterexamples, and thus grants to his empirical claim something like the status of an a priori truth.

To begin evaluating this objection we can rule out one extreme and unreasonable version of it. It would, in general, be an impossibly high standard to demand of disputants that every otherwise well-founded empirical claim on which they rely be abandoned in the face of just any purported counterexample. This would be an especially egregious standard in the cases where the empirical claim is supported by a wealth of *uncontroversial* evidence, and the purported counterexample is a particularly *controversial* one. Any empirical claim with a wealth of evidence supporting it should at least enjoy some slight benefit of the doubt in the face of merely *prima facie* counterexamples. The correct procedure in such cases would seem to be to undertake a careful scrutiny of the purported counterexample to see, at least, whether it could be explained in some way that did not contradict the otherwise well-founded empirical claim, and perhaps in doing so to take into account what would be gained or lost were the

counterexample to prove to be genuine. It is this very reasonable procedure, and not the rash brandishing of the Copy Principle as an irrefutable a priori principle, that Hume does undertake when confronted with one of the Copy Principle's most formidable purported counterexamples, the idea of necessary connection:

Shall the despair of success make me assert, that I am here possess'd of an idea, which is not preceded by any similar impression? This wou'd be too strong a proof of levity and inconstancy; since the contrary principle has been already so firmly establish'd, as to admit of no farther doubt; at least, till we have more fully examin'd the present difficulty. (T 1.3.2.12; SBN 77)

Hume here appeals to the fact that the Copy Principle is supported by a great deal of evidence in order to warrant, not tossing aside the purported counterexample, but rather investigating the matter further.¹⁶ Of course, what Hume finds in these further investigations is directly relevant to whether or not he ought to consider such purported counterexamples genuine. To stick with the current example—necessary connection—Hume finds that this idea would not only be a counterexample to the Copy Principle, but would also have to be utterly unlike any other ideas with which we find ourselves. That is, it couldn't be just some mental entity intuitively like an idea that snuck in without being copied from any impression—like the missing shade of blue;¹⁷ rather, it would have to be a wholly different *kind* of mental entity, which would in turn require a complete overhauling of an otherwise well-tested *system* of explanatory hypotheses. None of this is to say that it is *impossible* that we have an idea of metaphysically

¹⁶This is a point that Don Garrett also makes in Garrett, (1997).

¹⁷As we will see momentarily, the missing shade of blue is an example of an idea that has not been copied from any impression. It is, however, also an idea that *might as well have been* copied from some idea. Hume has a story to tell about how the idea of the missing shade of blue could arise from impressions of neighboring shades of blue. Part of Hume's point about necessary connection is that no such story could possibly be told about it.

necessary connection, but it is certainly enough to warrant proceeding to explore alternative avenues.

Furthermore, there is in Hume's favor the very fact that he is able to plausibly use the Copy Principle to ground the intuitive distinction between impressions and ideas. Anyone who wanted to claim that there are some ideas that are not copied from impressions would then be left with the burden of explaining just what distinction they were employing in making such a claim. We have already seen the difficulties attending such a project, and it is certainly to Hume's credit that he is able to marshal such a well-evidenced empirical claim to do this work.

All of which is not to say that Hume might not be wrong in the end. Rather, the point has been merely to show that there is a way that Hume can use the Copy Principle to ground the distinction between impressions and ideas without having to treat it as either an analytic truth about mental entities, or a mere empirical claim. It does have an empirical component, but it is not a *mere* empirical claim. It is a well-founded empirical claim that, because we have some independent *intuitive* grasp on the distinction at hand, can be used as a *theoretical underpinning* of that distinction.

Now, to tidy things up, we must turn to the missing shade of blue example. The case is as follows. Suppose that a person has had impressions of, and formed ideas of, a wide variety of shades of blue. Suppose, in fact, that—assigning each shade of blue of which they have had an impression a letter—they have impressions of shades A through L and N through Z. Hume admits that they would, though they have never had an *impression* of M, be able to form an idea of it. The first thing to notice about this example is that it differs from, say, the necessary connection example in that it is an instance of a

non-controversial purported counterexample, so its standing as a potential counterexample to the Copy Principle is immediately *prima facie* better.

The second thing to notice about it is that once again it relies on our pre-theoretical, intuitive notion of what the distinction is between impressions and ideas. That is, if what it *means* to be an idea is just to be a mental entity that is a copy of some other mental entity, this case should be impossible. The first mental entity that is of this M should be considered an impression of it. Intuitively, however, this seems wrong. The perception of M—perhaps because of its degree of force and vivacity, perhaps because of the context in which it appears, etc.—just seems more plausibly regarded as an idea, more like a memory than like a first encounter. Given that, however, this seems like a powerful counterexample not only to the truth of the empirical component of the Copy Principle, but also to Hume’s attempt to use it to ground the intuitive impression/idea distinction.

The position that I will now defend is that the Copy Principle *qua* empirical claim *does* admit of this exception, as Hume explicitly tells us, but that this does not undermine the use to which he puts it in drawing the impression/idea distinction. The difficulty with the latter part of this claim is that if what it is to be an idea is to be a copy of some mental entity, and what it is to be an impression is to *not* be a copy of any other mental entity, then it seems that admitting that there could be, and actually *is*, an idea which is not a copy of some other mental entity seems to make no sense. The way out of this predicament, as we are about to see, is to recognize that the idea of M resembles the impressions that caused it in such a way that it is *as good as* a copy of those impressions.

It will be helpful to beat around the neighboring bushes here, starting by contrasting this idea of M with impressions of reflection—another class of mental entities that differ from impressions of sensation in that, while not *copies* of other mental entities, they are at least *caused by* other mental entities. What earns impressions of reflection the appellation “impression” is that while they are *caused by* other mental entities, they are not *copies* of other mental entities because they do not resemble (exactly or roughly) the mental entities that cause them. Thus their circumstances, or essential qualities, are original to the mind, even though their causal history is not.¹⁸ The idea of M seems to share these features; it is caused by other mental entities (impressions and ideas of A-L and N-Z), and yet it does not exactly resemble any of these. The important difference between impressions of reflection and the idea of M, however, is that while the circumstances of impressions of reflection are original to the mind, the story is a little more complicated with the idea of M. Its circumstances do not *exactly* resemble that of the impressions that caused it, but do resemble them *in a way*. This phenomenon is also displayed elsewhere in Hume’s system. *Complex* ideas often do not exactly resemble the impressions that cause them, but because they are composed of simple ideas that do

¹⁸It follows from our construal of the impression/idea distinction that if we are to have an *idea* of an impression of reflection—say of anger—then that idea will have to exactly resemble the impression and be caused by it. Certainly, the second of these conditions is unproblematic—ideas *are* caused by the impressions of which they are copies on Hume’s account—but the second condition might be thought to be more troublesome. It would seem to follow from our account that any idea of anger that we form itself *has* all the properties that our actual *feeling* (impression) of anger has. Our idea of anger, that is, must itself be an *instance* of anger as well. To answer the above worry, then, ideas of impressions of reflection *do* share the essential characteristics of those impressions. Such must be admitted by *any* strictly Humean account of impressions and ideas. (For instance, the force and vivacity account will say that these share their intrinsic characteristics, and only differ in their degree of force and vivacity.) Furthermore, any strictly Humean account will also have to admit that the only differences between such an idea and its corresponding impression are differences in force and vivacity, and differences in relational properties. The only difference between the current interpretation and, say, the force and vivacity interpretation is that the current one takes these claims that Hume *clearly* takes as *at least* truths about impressions and ideas, and treats them also as playing a theoretical explanatory, and thus, criterial role in constituting the distinction between such perceptions. Such an objection, insofar as it is damaging, is equally damaging to all construals of Hume, as it is an objection to something that Hume *clearly* takes to be a true thesis of his.

exactly resemble the impressions that cause *them*, they are properly classifiable as ideas. Of course, the idea of M is, presumably, a *simple* idea, so we cannot explain its status as an idea as deriving from its component parts.

Now, while the idea of M is like a complex idea in that it does not *exactly* resemble the impressions that caused it, and it is different from a complex idea in that it is not composed of *parts* that exactly resemble the impressions that caused it, it does still resemble these impressions *in some sense*. Specifically, it resembles them in a very particular instance of the way that colors and other simple ideas can resemble each other. Hume writes of this more general phenomenon that,

‘Tis evident, that even different simple ideas may have a similarity or resemblance to each other; nor is it necessary, that the point or circumstance of resemblance shou’d be distinct or separable from that in which they differ. *Blue* and *green* are different simple ideas, but are more resembling than *blue* and *scarlet*; tho’ their perfect simplicity excludes all possibility of separation or distinction. (T 1.1.7.7n; SBN 637)

Clearly, Hume thinks that simple ideas, and specifically colors, can resemble each other. The missing shade of blue is a very particular instance of this kind of resembling. The idea of M resembles the ideas of L and N *so much* that we can arrive at M simply by filling in the gap between L and N. The very fact of our being able to do this, in turn, shows that the resemblance is so close between M and N, and M and L, that while the idea of M is not strictly speaking copied it is *as good as* copied. It is not as if, as is the case with impressions, the content of M is *utterly* inexplicable. The example, in fact, is only plausible because it is so obvious where the content of M comes from: L and N. So, while the Copy Principle does admit of this exception, it can still ground the impression/idea distinction because this is not an example of, so to speak, a *completely* original idea, which would undermine this use, but rather it is an example of a perception

that is *as good as* a copy, and so properly classifiable as an idea. To see this we need only note that had Hume chosen to change the Exact Resemblance Condition of the Copy Principle to something like, “A copy must exactly resemble that which caused it, *or resemble that which caused it in the way that M resembles L and N,*” not much would be lost in his subsequent use of it.

Before we move on, there is an important dialectical objection that we must now consider. One might want to object here that given the line that we earlier took towards the force and vivacity reading, such an account of what goes on in these passages ought to be off limits to us now. That is, earlier we objected to the force and vivacity reading of the distinction between impressions and ideas on the grounds that there are places in his writings where Hume clearly relies on something *other* than force and vivacity to draw this distinction. If, however, it is part of our own reading that at times Hume relies on our pre-theoretical, intuitive ability to make this distinction, it would seem that the proponents of the force and vivacity reading ought to be able to appeal to this pre-theoretical use in the cases that were cited as problematic for them. That is, the proponents of the force and vivacity reading might account for the instances in which Hume relies on some criterion other than force and vivacity by appealing to Hume’s use of the pre-theoretical, intuitive notion of the distinction between impressions and ideas just as we have.

Furthermore, it should be noted that while Hume relies on our pre-theoretical notion of force and vivacity to motivate the intuition that the idea of the missing shade of blue is an *idea*, nothing in *this* procedure requires him to refer to extra-mental entities. Hume does not, and need not, present the case as one in which a person has encountered

external objects A-L and N-Z and somehow comes by the idea M. The case is of a person who has had such-and-such impressions, *classified according to the criterial use of the Copy Principle*, not according to external causes of such impressions. Doing otherwise would violate Hume's own sanction on hypothesizing anything about the source of such impressions, and would be wholly unnecessary.

The problem with this objection is that it presupposes that the force and vivacity reading and the Copy Principle reading interpret Hume as being up to the *same kind* of activity, but employing different means of achieving it, but this is not the case. That is, in order to make this objection, the proponents of the force and vivacity reading would have to portray Hume as using force and vivacity to provide an *explanatory account* of our pre-theoretical notion of the distinction between impressions and ideas. The problem with their doing this is that force and vivacity are clearly *part of* this pre-theoretical, intuitive picture, whereas the Copy Principle is not. We can explain Hume's use of some criteria other than the Copy Principle by pointing to the pre-theoretical, intuitive picture. The proponents of the force and vivacity reading, first of all, have nothing specific to point to at all (as we can point to force and vivacity), and, secondly, have no explanation of why there ought *to be* anything other than force and vivacity that might be used in this way.

Now that we are fairly clear on what the Copy Principle is, and the sense in which it grounds the distinction between impressions and ideas, our next task is to return to those bits of text that seemed to support the force and vivacity reading, and see if we can account for them in some other way. We have already encountered a paradigm example of the most important class of such texts. It is the following.

All the perceptions of the mind are of two kinds, *viz.* impressions and ideas, which differ from each other only in their different degrees of force and vivacity. (THN 1.3.7.5; SBN 96)

Remember that the problem here is supposed to be that if it is literally true that the *only* difference between impressions and ideas is their degree of force and vivacity, then the only *candidate* available for that which makes a mental entity an impression or an idea is degree of force and vivacity; nothing else could possibly do the job. The key here, of course, is to see that this superficial reading of the passage is *not* literally true: that the only difference between impressions and ideas is their degree of force and vivacity. This much is not controversial. Hume clearly holds that the Causal Condition of the Copy Principle is *true*, even if he does not take it that that principle grounds the distinction between impressions and ideas. So, at the very least, impressions, *as a matter of fact*, differ from ideas in their causal origins. If, however, they differ in this way, then the above quotation ought not to be taken as ruling out this difference as a candidate for what grounds the distinction between them. Rather, we ought to read Hume as here working with a something like a bounded quantifier. The “only” in this passage is not meant to range over *all* the qualities that a perception might have, but only the *non-relational* non-essential ones. So we ought to read this passage, and ones like it, as claiming that ideas and impressions (considered individually, or non-relationally) differ only in their different degrees of force and vivacity. Of course they differ in other ways, but all of these other ways are relational differences, which are not what Hume is focusing on in such passages.

The next class of texts that seems to need explanation is one in which Hume appears to draw the impression/idea distinction in a way that is clearly other than via the

Copy Principle. These are a kind of mirror to the texts that we earlier presented by way of objection to the force and vivacity reading. Remember that there we presented texts in which Hume explicitly writes that the degree of force and vivacity of an idea could be equal to that of an impression. The parallel text here would be one in which Hume explicitly writes that there could be some idea that was not a copy of any other mental entity. We have, of course, just encountered such a class of texts: those concerning the missing shade of blue, in which Hume is *defending* the Copy Principle. On our construal, however, these texts do not lend support to the force and vivacity reading, because they do not in fact undermine Hume's use of the Copy Principle to ground the distinction between impressions and ideas. Remember that according to our reading of these texts, this is a distinction that we can intuitively make—via symptomatic features of the difference (force and vivacity), or paradigm examples, etc.—but which is not *defined* by any of these. The Copy Principle comes into play as first, an empirical fact about the mental entities so classified, and then as that which actually underlies this distinction. This entitles Hume to say, as he does, both that as a matter of empirical fact all ideas are copies of impressions, and that, as it turns out, impressions and ideas just *are* originals and copies, respectively.

By way of concluding, I will delineate some of the advantages that the reading of Hume's impression/idea distinction so far presented enjoys. Firstly, as we have already seen, reading Hume's distinction in this way enables us to make sense of the passages in his works in which he is clearly using some criterion other than force and vivacity to distinguish impressions from ideas, and I contend that this exegetical gain does not have any correspondingly significant exegetical cost. Secondly, this reading slightly lightens

the burden that Hume has traditionally been seen as laying on force and vivacity. That is, Hume uses degree of force and vivacity in other places in his system to account for various other mental phenomena (e.g., belief, time order, etc.) and the more work this has to do the less plausible it is that it can do it all. Therefore, by reading the impression/idea distinction using the Copy Principle we free up degree of force and vivacity to do more work elsewhere.

Lastly, there is a large body of contemporary literature surrounding what might be called causal theories of conceptual content (think Dretske, Fodor, Millikan, etc). Reading Hume's distinction using the two-part Copy Principle it becomes quite natural also to use this principle to fix the content of impressions and ideas.¹⁹ (This line would start with the rudimentary claim that an impression or idea is *about* that which in its causal history it exactly resembles.) This, in turn, places Hume's system right alongside contemporary participants in this debate as another systematic account/research program to be reckoned with—with its attendant advantages (e.g., resemblance as a solution to so-called disjunction problems) and disadvantages (having to cash out the notion of resemblance, etc.). Of course, as we noted earlier, Hume is not one for the external world, so on *strictly* Humean principles it would turn out—taking this line—that ideas are all and only of impressions, and that impressions have no intentional content. This seems to be both a thoroughly Humean line to take on such things, but also a bit troubling. The unsettling feeling of this line, however, is—one supposes—of a piece with that of denying the sense in talk of an external world. Thus is the price of being a *strict* Humean; one can always be less strict.

¹⁹Cf. Fodor, (2003).

CHAPTER TWO

A (Sellarsian) Kantian Critique of Hume's Theory of Concepts

One of the many revolutionary changes that Kant instigates in his *Critique of Pure Reason*, and perhaps the most important of these changes for contemporary philosophy of mind, is his presentation of his theory of concepts according to which a concept is a *rule for judging*.²⁰ The details of this theory are not the primary concern of this paper; Kant's reasons for preferring it to the other theories of concepts available at the time are. Of course, some of those reasons are deeply connected to the most difficult parts of Kant's system: the Transcendental Deduction, the transcendental unity of apperception, the unity of theoretical and practical reason, etc. None of these is the focus of this paper either. What I will concentrate on is a particular line of objection that Kant pursues in arguing against Hume's theory of concepts (and that contemporary Kantians pursue in arguing against contemporary Humeans) having to do with the logical form of judgments.

Hume's theory of concepts has as its foundation his theory of the mental, according to which everything mental can be accounted for in terms of mental entities

²⁰Kant, (1998).

and their relations and behaviors. He calls such entities perceptions, and divides these into two kinds: impressions and ideas. It is ideas that are involved in explaining the content of concepts and judgments, for Hume, and so it is ideas that we will focus on here. Content, for Hume, first accrues to mental entities in their role as singular representations. According to Hume, the content of an idea, what that idea is an idea *of*, is determined by his two part Copy Principle. For one entity to be a copy of another entity, on Hume's account, the former must both be caused by, and exactly resemble, the latter. Ideas stand in just this relation to impressions, and are therefore copies of them. On Hume's (imagistic) account, this also means that the content of any idea is just the impression of which it is a copy.²¹ It is the fact that ideas exactly resemble and are caused by impressions (or objects when the chips are not yet all down) that makes them *about* such things.

The next stage in Hume's project is to give an account of the content of general representations—abstract ideas, or concepts—in terms of the content of singular ones. Hume accepts Berkeley's famous rejection of Locke's account of abstract ideas (as indeterminate representations) on the grounds that ideas are mental entities, and everything that exists is fully determinate. So, given Hume's commitment to explaining the mental in terms of mental entities, their behaviors, and their relations, it has to be something about the relations and behaviors of ideas that explains our ability to think abstractly. Indeed it is. On Hume's account, a general representation is just a singular idea playing a particular role in thinking. Namely, a singular idea serves as an abstract

²¹This is Hume's view when all the chips are down and he is being a skeptic about the external world. When, however, the external world is back in play, there exists a causal chain from objects, through impressions, to ideas, and since these three entities will be the only nodes along the causal chain that exactly resemble one another, Hume can account for the content of ideas, using the Copy Principle, to single out objects, impressions, or both.

idea when it is associated with a word that triggers the mind to summon forth ideas that resemble this idea. Again, such a singular idea becomes associated with a certain word in public discourse in such a way that a certain custom is established such that this idea is summoned by the mind when that word is encountered along with ideas that resemble it (in the appropriate ways—an important and dangerous clause, which we will simply allow Hume). So, for instance, on this account, our concept of red is a particular idea of red that is brought about by encountering the word ‘red’, which encounter also triggers the mind to call forth all and only its other ideas of red things. These ideas of red things are the content of the abstract idea of red.²²

With Hume’s account of the content of singular and abstract ideas thus under our belts we can now move on to Hume’s theory of judgment. Hume presents this theory as an account of *belief*, and its core is the claim that a belief is nothing more than an idea with a particularly high degree of force and vivacity. Hume explicitly contrasts this theory of belief with those according to which judgment consists in the uniting of ideas.²³ He explains that his account is a reductive analysis of notions like conceiving, judging, and reasoning, according to which,

What we may in general affirm concerning these three acts of the understanding is, that taking them in a proper light, they all resolve themselves into the first, and are nothing but particular ways of conceiving our objects. [...] the act of the mind exceeds not a simple conception (T 1.3.7.5n; SBN 97)

Hume’s project, as he conceives it, is to account for all judgment (and all reasoning) in terms of single conceptions properly enlivened. A belief, or a judgment, is not a *uniting*

²²THN 1.1.7.7; SBN 20-1.

²³Hume is first and foremost concerned to account for the distinction between merely entertaining and believing. Drawing *this* distinction, however, requires him to touch in important ways on the relation of ideas to judgments, if only to say *what it is* that one merely entertains or believes. This, in turn, allows us to draw on his discussion here as a source of further theory.

of ideas, according to Hume, but rather is a single idea properly enlivened. Hume does not give the details of how this reduction is supposed to be carried out, and it is Kant's contention that this is a case where the devil is indeed in those details—as we will now see.

The most powerful criticism that a Kantian can bring to bear on Hume's theory of concepts and judgments, taken in isolation from other parts of Hume's system, involves what Kant calls the logical forms of judgment. In its most basic form, this is a challenge to Hume to cash the promissory note that we saw him issue for the details of the story of how to reduce judgments to conceptions. That is, Hume takes as his basic unit of thought the idea, and claims that all forms of thought can be reduced to modifications of, relations between, and behaviors of, ideas; this Kantian objection asks to see how this is to be done, and along the way shows that each option available to Hume runs into serious difficulties. These difficulties are then diagnosed as being the inevitable result of a misbegotten project, and the Kantian then proposes his own theory of concepts and judgments as a viable alternative. We will implement the first step in this procedure here by examining what resources a Humean has available to complete this project, and evaluating the success of the various lines the Humean might take as we go.

To begin, consider the judgment 'this sphere is red'. It is a consequence of Hume's theory of abstract ideas, sketched above, that a sphere is red iff the idea of the sphere belongs to the set of mental entities that are, or would be, summoned according to the correct custom upon encountering some manifestation of the word 'red'. Abbreviating, we can say that for a sphere to be red, the idea of that sphere must belong

to the Red Revival Set²⁴ (where the use of capital letters signals that we are not speaking of a revival set that is itself red, but rather the revival set appropriate to the use of the word ‘red’).²⁵ Knowing, however, what it is for a sphere to *be* red is not yet to have an explanation of what it is to *judge* (and think, say, believe, et al.) that it is red. What we are looking for, what Hume needs to be able to provide to meet the Kantian challenge, is an account of the mechanism by which judgments *come to* mean what they do.²⁶ Of course, we know *that* (and in most cases *what*) they mean; this is how we are able to understand and evaluate Hume’s claim regarding what makes them true. What we need is an account, from Hume, of *how* they mean. This is what the Kantian thinks Hume cannot provide.

Consider again, then, the judgment ‘this sphere is red’. Given that Hume’s project is meant to dissuade us from the idea that a judgment is a uniting of ideas, the most straightforward way of explaining this judgment is as consisting of a *single* idea: that of the red sphere. On this account, to judge that the sphere is red is just to have the idea of the sphere, which will also, as it turns out, be an idea of something red. The problem with going this route is that since all ideas are fully determinate in all of their qualities, the idea of the *red* sphere will equally be an idea of, say, a sphere that smells like corn, a sphere that is brightly colored, a sphere that is smooth, etc. If, however, all judgments are reducible to single ideas as proposed, all judgments would be indeterminate with regard

²⁴I am borrowing this term from Garrett, (1997).

²⁵We can, for the purposes of this paper, grant Hume free use of sets, classes, etc. It is worth noting, however, that such a use is something that needs to be earned, and that it is unclear whether Hume has actually done this. (Certainly, e.g., Kant and Hegel would object that he has not.)

²⁶As we will see, part of Hume’s answer to this challenge will be to transform the question of how a judgment comes to mean what it does into the question of how an idea comes to serve as this or that judgment.

to what quality is being predicated of some object. Single ideas are simply too rich, and therefore too ambiguous, to be, by themselves, judgments.

What is needed is a way of singling out some quality of a thing as the quality that the thing is being judged as having. This makes *abstract* ideas a particularly appealing ingredient for judgment, since this is exactly their function. So, another way that Hume might account for a judgment such as ‘this sphere is red’ is by casting it as a complex idea, consisting of the complex idea of the sphere and the complex idea that consists of the Red Revival Set (which consists of all and only ideas of red things).²⁷ This proposal both does justice to Hume’s claim that judgments can be reduced to single ideas (single “doubly” complex ideas, that is), and to the intuition that judgments in some sense *unite* ideas.

The way in which complex ideas unite ideas, though, is *not* the way in which judgments unite ideas.²⁸ To see this we can turn our attention briefly to a twentieth-century analog of this dialectic. In its contemporary incarnation, it is the linguistic version of Bradley’s Regress that Wilfrid Sellars presents as an objection to the use of a Platonic metaphysics of abstract entities to account for the meaning of predicate terms.²⁹ (We will draw the parallels to Hume and Kant once the argument is on the table.) The

²⁷It should be noted here that we are simply *granting* Hume the unity of complex ideas. This is because we are after bigger game—judgments—but similar points can be made to show that Hume has not even earned the former unity. Fodor, for instance, uses the complex concept ANTI(ANTI-MISSILE) to show that even complex concepts have a structure over and above mere association that contributes to the determinate meaning that they have. Fodor, (2003). For our purposes, however, we can set this issue aside, and concentrate just on the missing account of the unity of judgments. That is, we are letting Hume have that a complex idea may be nothing more than a series of simple ideas associated by resemblance, contiguity, and causation relations, despite misgivings we might have about this, in order to focus our attention on the specific issues of judgments instead.

²⁸The difference is essentially captured in the Kantian slogan “a complex idea is not yet an idea of a complex”.

²⁹Sellars, (1963b): 7-26.

objection, as Sellars formulates it, is to the claim that predicative expressions come by the meaning they have by *referring* to abstract entities (or, worse still, that *all* expressions come by their meaning in this way). For current purposes, it will be terminologically helpful to call this kind of referring expression a name—even at the risk of blurring whatever fine-grained distinctions there may be between these two terms in ordinary use.

The first step in the objection to this account of meaning is to notice that whatever a judgment is, it is at least more than a mere list of names.

(1) Joan, Judy, Jeffrey, Jessica

is not a judgment. Analogously, then, neither is

(2) ‘the sphere’, ‘is red’,

or worse still,

(3) ‘the’, ‘sphere’, ‘is’, ‘red’.

To make even clearer that (2) and (3) are not judgments, but mere lists, we can change the order in which the items on the list are placed.

(4) ‘is red’, ‘the sphere’

and

(5) ‘is’, ‘sphere’, ‘the’, ‘red’

now bear almost no resemblance to anything that we might mistake for a judgment.

These mere lists of names are quite obviously *not* judgments, and simply reordering such lists so that they superficially resemble recognizable judgments cannot make them into such. The point here is that the Platonist’s move to account for the meaning of predicate terms as names of abstract entities leaves us in need of an account of how such names

come together with names of objects to form judgments (rather than mere lists of names).³⁰

The next step that Sellars takes against the Platonist is to point out that it is of no help to him to posit that judgments differ from mere lists by being implicitly of the form,

(6) the sphere *exemplifies* redness

where ‘the sphere’ is the name of some sphere, ‘redness’ is the name of some abstract entity, and ‘exemplifies’ somehow ties the two names together. In this case, either ‘exemplifies’ is the name of some relation, in which case it is merely another name on the list—i.e., the purported judgment actually has the form,

(7) ‘the sphere’, ‘exemplifies’, ‘redness’,

which is obviously of no help, and gets the eponymous regress on its way, or ‘exemplifies’ is not the name of anything, in which case it is a meaningful term that is not a name, and the Platonist needs to explain why ‘exemplifies’ can receive the alternative treatment it does, whereas ‘red’ cannot.³¹

Applying this form of argument to the suggested Humean approach to judgments, we get an argument that merely having a complex idea is not yet to judge anything. It is,

³⁰At first, it looks like this is exactly the problem that Frege’s notion of “unsaturated concepts” is meant to solve, and so it may be. The idea behind this notion is that if the problem is that we cannot find a way to unite a list of names into a judgment—actually, this is not the *only* problem that unsaturated concepts are meant to solve, but it is the one that concerns us—then perhaps one of the members in this list is not, in fact, a name. Concepts are not names on Frege’s view; they are *unsaturated*. The problem with this solution, however, is that the *reason* concepts are not names is because names refer to objects, whereas concepts refer to *non*-objects. Even granting Frege this reference to mysterious non-objects, it should be obvious that this is of little help. We are still left with a list of *referring* expressions—names and pseudo-names—and no way to unite them.

³¹Lastly, Sellars offers his alternative account according to which it is that an ‘is red’ is to the right of a ‘this sphere’ that says that this sphere is red (configurations of particulars, while constituted by particulars, are not reducible to mere lists of particulars), and the meaning of ‘is red’ is explained in terms of its inferential role. This, however, is not the place to delve into the details of Sellars’ story (even though it bears a striking resemblance to Kant’s, and not coincidentally). Suffice it to say that it is by making the judgment the fundamental semantic building block, rather than the concept (or the intuition), that Kant is able to avoid the problem we are here raising for Hume.

rather, the Humean mental analog of having a list of names: a list of ideas. Each idea is contentful independently of being combined into a judgment just as in the Platonist's picture each name is independently contentful. That ideas derive their content from being copies of impressions (resembling the impressions that cause them) is just a way of filling out the story of what the reference relation is that the Platonist appeals to in his account of naming. That ideas refer to impressions (and/or objects when all the chips aren't yet down) and the Platonist's names refer to abstract entities makes no difference in this context. The Humean is still left with a list of the form

(8) 'This sphere', 'Red Revival Set'

rather than a judgment. So long as the elements of a judgment are conceived as entities that are contentful independently of their role in judgment, the problematic of how these come together to make judgments remains.

One might be inclined to respond here that is of crucial significance that on Hume's picture complex ideas are *held together* by relations of resemblance, contiguity, and causation, and that somehow these relations provide just the right kind of glue to hold ideas together in just the right way so as to make the resulting complex idea into a judgment.³² This move, an appeal to the circumstances under which mental lists are formed, however, will not work. Its analog in the Sellars case would be an appeal to the fact that the list of names 'Joan, Judy, Jeffrey, Jessica' is held together by the resemblance of the first letter of each name, or the inclusion of all of these names within one set of single-quotes, or the fact that these are the names of four members of my mother's immediate family, and thus one causes me to think of the next. Clearly, though,

³²As noted earlier, it is not clear that these relations can even hold together complex ideas in the way they need to. See fn. iii. We are, for the time being, giving Hume this much.

none of these facts, although all true, are enough to change the list of names into a judgment.

It should also be clear that separating the complex idea into a three-part idea consisting of an idea of the sphere, an idea of red, and a very special unifying idea of *is*, will also be of no help. This is just the move that the Platonist makes that gets him caught in Sellars' fatal dilemma. It would leave us in need of an account of why the complex idea consisting of the three ideas 'the sphere', 'is' and 'red' is more than a mere list of names.

Perhaps, though, *adding* something to the concepts with which we are working (an additional concept, an additional non-concept, etc.) is not what is needed. Perhaps *modifying* the concept at hand in some way would do the work both more efficiently and effectively. This is a strategy that Hume adopts in the slightly different context of explaining the 'is', not of predication, but of existence. Hume cites the example of the judgment 'God exists' as an instance of a judgment that can and ought to be accounted for in terms of a *single*-conception model precisely because belief in existence does not consist in adding a further idea to the idea of what is believed in, but rather consists in a *modification* of this idea (namely, a higher degree of force and vivacity).³³ Transposing this approach to the current context, the suggestion would be that a judgment is a *particularly forceful and vivacious* complex idea. This, again, will not do at all. A particularly forceful and vivacious complex idea is, on Hume's account, a belief in the existence of what that idea represents (as the idea represents it), and *that* is not what we are looking for. In the case of the current analysis of 'this sphere is red' that would amount to a belief in the existence of the complex idea consisting of the idea of the

³³THN 1.3.7.5n; SBN 96.

sphere and the idea of the Red Revival Set. I.e., it would consist in a belief in the existence of the sphere and the existence of all red things. Such a belief, I hope it is clear, is not a belief *that* the sphere is red.

If there is a lesson to be drawn from the set of considerations surrounding Bradley's Regress it is that structure matters. A list of names is not a judgment, a conjunction of mental entities is not a belief, a complex of ideas is not an idea of a complex, a configuration of particulars is more than a mere collection of particulars, etc. If one is to build something with a complex structure out of constituent parts, one *must* have a story to tell about what the structure of what is built is, over and above the mere collection of these parts. As we have seen, one way to do this is, instead of starting with just particulars, to start with, e.g., particulars *configured* a certain way.³⁴ As we have seen, complex ideas are not by themselves of the proper kind of configuration to do the work that they need to in judgments; they are mere conjunctions of simple ideas. This might, however, tempt one to propose on Hume's behalf that ideas can be *of* not only particulars, and collections of particulars, but also *of* facts. In that case, the judgment 'this sphere is red' could be accounted for as an idea of *the fact that* this sphere is red.³⁵

While this suggestion clearly represents an intriguing line of investigation, it is one that we do not have the space to explore here. It is simply too far astray from Hume's actual project to consider under the heading of a suggestion of how *Hume* might have cashed *his* promissory note. Hume is clearly committed to ideas being only of particulars;

³⁴Sellars starts with configurations of particulars. Wittgenstein, in the *Tractatus*, starts with facts. The difference is crucial to certain problems in ontology, but not particularly important to *us* just now.

³⁵John McDowell endorses something very much like this picture in *Mind and World*. Combine this with the ensuing force and vivacity approach to judgments, and something remarkably like Descartes in the Fourth Meditation pops out.

facts are just not part of his ontology. Of course, one consequence of this is that we are already in a position to see that Hume is in a bind. If all he has to work with is particulars, and the meaningfulness of ‘this sphere is red’ must be understood as the meaningfulness of a fact, and facts have a structure over and above being a mere collection of particulars, it becomes difficult to see how that fact can ever be represented by just a collection of particulars. It seems as if whatever represents facts, because facts are structured, is also going to have to be structured, and Hume simply does not have the resources to account for such a structure. This, however, is still a premature conclusion. There are still a number of options—for explaining, and explaining away this structure—that we have not yet explored, and it is to the next of these that we will now turn.

The thinking behind the first proposal that we considered was that since there seem to be at least two ideas contained in the judgment ‘This sphere is red’, ‘this sphere’ and ‘red’, maybe the best way to account for that judgment is by combining those two ideas in the only way Hume can: by making them into a complex idea. The problem with this proposal was that this complex idea failed to capture the way in which these ideas are in fact combined in that judgment. This next proposal tries to capture that. Noticing that what it is for the sphere to be red is just for it to be in the Red Revival Set, one might propose that what it is to judge ‘this sphere is red’ is just to have the idea (with the proper high degree of force and vivacity appropriate to a belief) of the Red Revival Set with the sphere *in* it, as one of its *constituents*. This proposal has the benefit of adhering to two of Hume’s most relevant claims in this area. The first is that judgments can be reduced to beliefs, which are just forceful and vivacious ideas; the second is that all belief (and

thereby judgment) is reducible to the modifications of *single* ideas, in this case the abstract idea that is the Red Revival Set.

Despite these exegetical advantages, however, the philosophical disadvantages of this proposal, at least in the rudimentary form so far presented, make it untenable. Consider the Red Revival Set. It consists of the set of all red things. The proposal here is that since the sphere in question is a member of this set, the judgment ‘this sphere is red’ might be reducible to an enlivened idea of this set. The problem is that if the entire set is enlivened, each member of that set is enlivened, and ‘this sphere is red’ becomes indistinguishable from ‘this fire-truck is red’, ‘this clown-nose is red’, etc. There is also the problem, encountered earlier, that the enlivening of the complex idea consisting of all red things already counts for Hume as a belief in the existence of all red things, which the judgment ‘this sphere is red’ is certainly neither identical to nor inseparable from.

Notice also that enlivening the idea of the sphere more than the other members of the Red Revival Set will not help. While this might serve to make the judgment more determinate with regard to which member of the set ‘red’ is being predicated of, it still fails to distinguish a judgment about the existence of the sphere from one about the sphere’s being red. It fails to distinguish a judgment whose content is that the sphere exists from one whose content considers the sphere *qua* member of the Red Revival Set. Furthermore, the more one pushes this line, that the idea of sphere and the idea of the Red Revival Set have different degrees of force and vivacity allotted to them, the more one makes this proposal like the previous one, in which the sphere and the Red Revival Set are two distinct ideas being considered in conjunction with one another. We have already seen why that proposal fails.

The difficulty here is that force and vivacity are simply doing too much work. If, that is, the enlivenment of an idea by force and vivacity are what make for a belief in the *existence* of that idea's content, this enlivenment is not also available to do the work of the 'qua' in thinking of the sphere qua a member of the Red Revival Set. This observation suggests a remedy to some of the current difficulties. The problem is that force and vivacity have too much work to do. One solution would be if some other non-relational extrinsic qualities of mental entities could be found to do this work. Suppose, for instance, that beliefs in the existence of the objects of ideas are accounted for in terms of degrees of force and vivacity, but predicative judgments involving modifications of revival sets were accounted for by some other quality, like juiciness. 'This sphere is red', the account would run, has as its content the Red Revival Set in which the idea of the sphere is made juicy, whereas the other members are not. Thus, the ideas 'this sphere' and 'is red' are united, in a sense, and yet the judgment is still about this sphere rather than a fire-truck, a clown-nose, etc.

The problem here is seeing what this proposal amounts to. Remember that Hume's theory here is meant to *explain* (or explain away) the phenomenon of judgment. Force and vivacity are able to play this role because our understanding of them is supposedly introspectively available and is thus available independently of Hume's theory. Juiciness, however, is not, according to Hume, something that we introspect. If we have any grasp of juiciness it is only as that which makes a judgment predicative, etc. Furthermore, if this strategy is to work, there *must be* a different introspectively available non-relational extrinsic quality of mental entities for *every* kind of judgment.³⁶ As an

³⁶Remember that Kant, for instance, lists twelve.

empirical claim, this is implausible. As an analysis of the meaning of judgments, it is ridiculous.³⁷

The final suggestion I want to consider is one that might already have occurred to some readers, especially given that we have been in Kant's neighborhood for some time now. Earlier we considered whether the judgment 'this sphere is red' could be accounted for by the fact that this sphere is *in* the Red Revival Set. Our final proposal is that cutting up these ideas a little more finely might do the trick. That is, instead of relying on *this sphere* being in *the Red Revival Set*, we ought to be focused on *this* being in *both* the Red Revival Set and the Sphere Revival Set. Perhaps, this suggestion runs, it is the idea's membership in intersecting revival sets that does the work of accounting for 'this sphere is red'.

Now, in evaluating this final suggestion, we have to be careful about a number of things. First of all, any Kantian worth his stripes will want immediately to object to the notion of non-conceptual thises on which this suggestion relies. That, however, is a deep and complex form of objection that will not concern us here. Secondly, we must keep in mind what we had cause to note at the outset of this paper: that it is one thing to be able to account for what it is for "this" to *be* red and spherical, and yet quite another to account for how we *judge* that it is red and spherical. It is not enough to cite, that is, "this" being a member of both the Red Revival Set and the Sphere Revival Set to explain how 'this sphere is red' comes to mean what it does. What Hume needs is an explanation

³⁷For lingering reasons connected with Bradley's Regress, for Wittgensteinian reasons involving publicity, etc. It should also be noted at this point that attempts at revitalizing this line of thought by having the idea of the sphere serve the purpose of *drawing one's attention* to its membership in the Red Revival Set will run into parallel difficulties as those proposals already considered.

of how *that fact* is represented by ‘this sphere is red’. That is the explanation that we will now see if Hume can provide.

With these potentially obscuring issues set aside, however, it ought to be clear that this final suggestion is no advance from where we started. What we have now is a list of meaningful terms—this, red, and sphere—which need somehow to be put together into a judgment. Separating the ‘this’ from the ‘red’ and the ‘sphere’ only makes it more difficult to see how to put such things back together again. With the failure of this final proposal, then, our exploration comes to an end. While this list most likely has not been exhaustive it certainly leaves a bleak picture in place for the future of Hume’s research project, and those of any other semanticist who embraces an atomistic view of conceptual content.³⁸

As I said at the outset, the purpose of this paper is, for the most part, to carry out the first two stages of the Kantian critique of Hume’s theory of concepts: to evaluate critically the details of the Humean’s promised reduction of judgments to concepts, showing each proposal inadequate to the task, and then to diagnose these successive failures as the result of a specific misbegotten project. We have now done this much. The third stage of that critique, which I was not to undertake, is to propose an alternative account of judgments and concepts that avoids these difficulties. I have no intention of

³⁸In fact, this need not be the end of this research project. The most fruitful move that contemporary Humeans have made is to pursue the thought that ideas are not exhausted by their content, but also have a form, and that it is in virtue of this form that they are combinable into judgments. Jerry Fodor, for instance, argues that *syntax* does exactly this job. There is not enough room here even to begin to explore such a line, but two points are worth noting. Firstly, such a line requires abandoning both Hume’s associationism and his theory of conceptual content—both theses eminently worthy of abandonment—and so can only be broadly construed as Humean insofar as it is still committed to the theses that the mental is explainable entirely in terms of mental entities and their relations and behaviors, and that meaning accrues first and foremost to concepts, not judgments. Secondly, there are reasons having to do with the crucial role that concepts and judgments play in Kant’s Transcendental Deduction of the Pure Concepts of the Understanding to think that even with such modifications, atomism simply will not do. Unfortunately, that is a topic for another paper.

renewing on my commitment to refrain from delving into that third stage, but it does seem appropriate at least to present a prospectus for how one might go about doing so, and that, therefore, is my plan for concluding this paper.

As I have mentioned throughout, Kant's solution to the problem of the unification of judgments is, in a sense, to avoid undertaking the commitments that make this a problem for Hume in the first place. Remember that the problem for Hume is that he is committed to accounting for everything mental in terms of mental entities and their relations and behaviors. Since concepts and judgments seem clearly to be (at least in part) mental phenomena, Hume's account of these must be made answerable to the above constraint. Additionally, Hume is committed to the thesis of conceptual atomism: simple ideas are the semantic building blocks from which all other contentful mental phenomena (such as judgments) are built.

Kant is fairly radical in rejecting both of these theses,³⁹ but for the purposes of this prospectus it is his rejection of the latter that will most concern us. Kant takes not the concept, but the *judgment* as the fundamental semantic building block. As he puts it, "the understanding can make no other use of [...] concepts than that of judging by means of them"(A68/B93). Concepts have the meaning they do *only insofar* as they play a certain role in the judgments that they do. E.g., the concept 'red' refers to red things not in virtue of standing in some relation to red things, but by playing the role it does in judgments such as 'this sphere is red', 'this red sphere is large', etc. Concepts are, so to speak, already infused with judgmental form. 'Red' is more perspicuously represented as '___ is

³⁹It is arguable that he is the first philosopher to reject either of these theses, although perhaps surprisingly, Plato gives him a run for his money on the rejection of conceptual atomism. Certainly he is the first Modern philosopher to do so.

red', 'this sphere' as 'this sphere...'.⁴⁰ It is the judgments in which concepts figure that gives these concepts their meaning, and so providing an account of the unity of judgments is unnecessary.

What *is* needed to make Kant's account complete is both a story about how judgments have the meaning that they do, and how concepts have the meaning that *they* do as a result of playing the role they do in judgments. That is, what we have just proposed is not yet anything like a *semantics* for either judgments or concepts. Instead it is a suggestion for what the starting place for such a semantics should be. As it happens, Kant does not tell a great deal of the full story of conceptual and judgmental content. This is exactly the project, however, that Hegel takes up in his *Science of Logic*, and it is the story of an *inferentialist* account of meaning.⁴¹ This, however, takes us as close as we are willing to trespass to the details of such a solution in the context of the current investigation, and so represents our end point.

⁴⁰This is, in certain respects, surprisingly similar to the solution to our problem that Jerry Fodor presents on Hume's behalf in *Hume Variations* (see fn. xv). In other important respects these accounts are *very* different.

⁴¹ Hegel, (1976).

Chapter Three

Inferentialism and the Transcendental Deduction

One recent trend in Kant scholarship has been to read Kant as undertaking a project in philosophical *semantics*, as opposed to, say, epistemology, or transcendental metaphysics.⁴² This trend has evolved almost concurrently with a debate in contemporary philosophy of mind about the nature of concepts and their content. Inferentialism is the view that the content of our concepts is essentially inferentially articulated, that is, that the content of a concept consists entirely or in essential part in the role that that concept plays in a system of inferences. By contrast, relationalism is the view that this content is fixed by a mental or linguistic item's standing in a certain relation to its object. The historical picture of Kant and the contemporary debate about concepts intersect insofar as contemporary inferentialists about conceptual content often cite Immanuel Kant not only as one of the founding fathers of a tradition that leads more or less straightforwardly to contemporary inferentialism, but also as the philosopher who first saw the fatal flaws in any attempt to articulate the content of our concepts relationally.⁴³ Kant's advances over his predecessors (e.g., Descartes, Locke, and Hume) are cited as examples of the triumph

⁴²Cf. Sellars, (1967), Brittan, (1978); Pippin, (1982); Brandt, (1995); Longuenesse, (1998); Hanna, (2001); Rosenberg, (2005).

⁴³Cf. Sellars, (1967): chapter 1. Robert Brandom correctly cites Kant's key inferentialist insight as concerning the primacy of the proposition Brandom, (1994): 79-80, but does not concern himself with Kant's *arguments*. Rosenberg, (2005), traces a line in Kant's thought to which the current study owes much.

of inferentialism over relationalism. On the other hand, contemporary relationalists maintain that it is only the meager resources that these philosophers misguidedly permitted themselves that allowed Kant to win the day. A more state-of-the-art relationalism, they argue, is more than adequate to meet the challenges laid down by Kant.⁴⁴

The purpose of the current paper is to revisit the historical debate—specifically between Kant and Hume—to see (a) whether Kant can plausibly be read as an inferentialist, (b) if so, what kind of inferentialist he is, (c) how Kant argues against Hume’s view, (d) whether these arguments are sound. I will argue that Kant *is* an inferentialist, that his arguments against Hume’s relationalism rest on certain further considerations that he marshals regarding the role of concepts in our mental lives, and that these considerations commit him to a very particular brand of inferentialism according to which what is represented by a concept is not an object or set of objects but a *way* that objects can be.

Hume sets out in *A Treatise of Human Understanding* to provide a complete scientific account of the mental lives of human beings. He begins there by drawing a distinction between two kinds of mental entities. ‘All the perceptions of the human mind resolve themselves into two distinct kinds [...] impressions and ideas’ (T 1.1.1.1; SBN 1).⁴⁵ He then further divides each of these classes of perceptions into those that are simple and those that are complex. ‘Simple perceptions or impressions and ideas are such as

⁴⁴The most vocal relationalist in this regard is Jerry Fodor, whose Fodor, (2003) explicitly undertakes to give an updated version of Humean relationalism.

⁴⁵Hume, David, *A Treatise of Human Nature* eds. David Fate Norton, Mary J. Norton. (Oxford: Clarendon Press, 2000) and Hume, David *A Treatise of Human Nature* ed. L. A. Selby-Bigge. (Oxford: Clarendon press, 1896). Quotations from the *Treatise* are cited in both the Norton edition (by paragraph number) and the Selby-Bigge edition (by page number).

admit of no distinction or separation. The complex are the contrary to these, and may be distinguish'd into parts'(T1.1.1.2; SBN 2). For Hume, the mind contains atomic impressions and ideas—simple perceptions—and perceptions that are built up from these—complex perceptions. Hume next observes that these simple ideas are caused by and exactly resemble their corresponding simple impressions, and concludes that all simple ideas are *copies* of simple impressions.⁴⁶ Since the causal condition and resemblance condition, respectively, jointly constitute what it is to *be* a copy for Hume, his conclusion straightforwardly follows from his observations.

Although Hume never explicitly announces that he is doing so, he goes on in the subsequent parts of the *Treatise* to employ this “copy principle” as a *semantic* principle, as determining the *content* of ideas.⁴⁷ That is, he goes on to suppose that the content of any given simple idea is just the simple impression from which it has been copied, i.e., which it exactly resembles and caused its existence.⁴⁸ Because complex ideas are not exact copies of complex impressions, and because they can be caused by complex impressions that they do not resemble⁴⁹, Hume’s semantic story about them is both more

⁴⁶It is worth noting that in observing that simple impressions are the cause of simple ideas, Hume does not employ the notion of causation that he shows to be problematic later in the *Treatise*—necessary connection—but rather appeals to exactly the alternate account of causation that he later proposes: constant conjunction and precedence.

⁴⁷Kemp, (2000) does a nice job of presenting a reading of Hume as expressly concerned with the *content* of mental items, and his use of the Copy Principle as content-fixing. For a brief discussion of how this principle plays a crucial role in some of Hume’s most important arguments cf. Landy (2009b).

⁴⁸It is worth noting that on this reading, for Hume, impressions have *no content* because they are not copied from anything. This goes some way towards explaining why Hume cites the difference between impressions and ideas as being, “the difference betwixt feeling and thinking”(T 1.1.1.1; SBN 1). Of course, Bennett, (1971) argues that because impressions *must* have content, they must be copied from objects in the external world. Everson, (1988) does a nice job of rebutting this suggestion, but see Landy, (2006) for a necessary corrective to the conclusion that Everson draws from his argument.

⁴⁹“I observe that many of our complex ideas never had impressions, that correspond to them, and that many of our complex impressions never are exactly copy'd in ideas”(T1.1.1.4; SBN 3).

complicated and little more obscure. The general idea is that a complex idea is *composed* of simple ideas each of which has its content via being a copy of the impression which is its object. Complex ideas, then, have as their content the aggregate content of their constituent simple ideas.⁵⁰

It is important to notice here that Hume's account of the content of simple and complex ideas is a form of relationalism. A simple idea has the content that it does by standing in a certain relation to the impression that it represents; it is a *copy* of it.⁵¹ Similarly, the content of a complex idea is an aggregate of the content of its constituent simple ideas, which is determined relationally. This relationalism about ideas, we are about to see, applies also to Hume's theory of abstract ideas, the closest thing in Hume's system to what Kant calls "concepts".

Hume summarizes his account of abstract ideas as follows.

When we have found a resemblance among several objects, that often occur to us, we apply the same name to all of them, whatever differences we may observe in the degrees of their quantity and quality, and whatever other differences may appear among them. After we have acquir'd a custom of this kind, the hearing of that name revives the idea of one of these objects, and makes the imagination conceive it with all its particular circumstances and proportions. But as the same word is suppos'd to have been frequently apply'd to other individuals, that are different in many respects from that idea, which is immediately present to the mind; the word not being able to revive the idea of all these individuals, only touches the soul, if I may be allow'd so to speak, and revives that custom, which we have acquir'd by surveying them. They are not really and in fact present to the mind, but only in power, nor do we draw them all out distinctly in the imagination, but keep ourselves in a readiness to survey any of them, as we may be prompted by a present design or necessity. (T

⁵⁰What one *wants* to say here, of course, is that a complex idea has as its content more than just the *aggregate* content of its constituent simple ideas; it has as its content some *structured* whole whose parts are the content of these simple ideas. Unfortunately, *Hume* cannot say this for reasons having to do with his theory of complex representation. Cf. Landy, (2009a).

⁵¹Remember that, strictly speaking, for Hume we do not have any idea of the external world. All our ideas are of mental items. Perhaps a less skeptical version of Hume could use the Copy Principle to fix worldly objects as the content of our ideas, if they held that our ideas are caused by and resemble such objects.

1.1.7.7; SBN 20-1)

Upon encountering a number of objects that all resemble one another, we come to call them all by the same name. Eventually we form a habit of associating this name with these objects, so that an encounter with the one produces an idea of the other. Still further on in the process, we come to abridge this association so that upon hearing a certain name, we no longer immediately call to mind *all* the objects with which that name is associated, but only some. These some stand for all the objects, however, insofar as we are disposed to recall those others, if the need were to arise—for instance, if we needed to draw conclusions about all the objects on the basis of conclusions that we draw about this some.⁵²

The single idea that is called to mind by the hearing of the word is “abstract” just insofar as it stands in the proper relation to the other ideas also associated with that word. The content, then, of this abstract idea is just the sum of the content of all of the ideas so associated with it. Here is Hume describing how a mind not yet comfortable with the use of some abstract idea, can familiarize itself with its content.

Before those habits have become entirely perfect, perhaps the mind may not be content with forming the idea of only one individual, but may run over several, in order to make itself comprehend its own meaning, and the compass of that collection, which it intends to express by the general term (T 1.1.7.10; SBN 22)

What the mind intends to express by a general term is the content of the collection of ideas that are associated with the general term. The content of one’s idea of, say, ‘elephant’ is just the sum of the content of all of the ideas that one associates with the

⁵²“Thus shou’d we mention the word *triangle*, and form the idea of a particular equilateral one to correspond to it, and shou’d we afterwards assert, *that the three angles of a triangle are equal to each other*, the other individuals of scalenum and isosceles, which we over-look’d at first, immediately crowd in upon us, and make us perceive the falsehood of this proposition, tho’ it be true with relation to that idea, which we had form’d”(T1.1.7.8; 21).

word 'elephant'. Again, there are nuances and difficulties here that need not concern us. What does concern us is the sense in which this is a *relationalist* theory of concepts.

At first blush, it might be difficult to see how Hume's account of abstract ideas is a relationalist one. The content of such ideas does not seem to be fixed by relating ideas to the objects of which they are copies, but rather it is fixed by relating ideas to *one another*. At the outset of this investigation, this seemed to be the very hallmark of a *non-relationalist* account, such as inferentialism. The key to seeing why contemporary relationalists nonetheless cite Hume as a forefather is to see that while his account of the content of concepts does *involve* the relations among ideas, it does so in a way that is fairly innocuous for the more general project of relationalism. In particular, it is an essential part of Hume's account of the content of abstract ideas that such ideas *do* relate to the objects of which they are copies; it is just that on that account, they do so *indirectly*. It is still the case that the content of an abstract idea, for Hume, is given by the relation that that idea bears to the object of which it is a copy. An abstract idea has as its content that which the ideas that are "annexed" to it have as *their* content, and these ideas have as their content that of which they are copies. An abstract idea, therefore, has the content that it does by standing in a certain relation to *collections* of copied objects, rather than single ones. Hume makes use of a principle of collection, which involves relations between ideas, but the content conferring relation remains essentially the same; it is the copy relation applied to collections rather than individuals. The annexed ideas are each copies of such-and-such objects, and the collection of them has as its content just the collection of their content. The key to relationalism is the delineation of the content of a

concept as essentially a word-world, or idea-object, relation.⁵³ Hume's account keeps this fundamental commitment intact; the content of an abstract idea is still a function of the relation of that idea, and those ideas that resemble it, to their objects.

Before we move on, it is important to note the representational work that such ideas perform in Hume's system. Hume, like most relationalists, takes it that the work that a concept does is similar to that of concrete ideas; it stands for something. Just as a concrete idea stands for a concrete particular, an abstract idea stands for—according to the specifics of Hume's account—a collection of such particulars. As Hume remarks regarding his theory of abstract ideas, it is sufficient for his purposes if he can show—without violating either of the commitments listed earlier—that his theory is able to account for the fact that,

Some ideas are particular in their nature, but general in their representation. (T 1.1.7.10; SBN 22)

What Hume wants to show is that we are able to represent a plurality of objects using just a single idea. As we will see in a moment, one of the crucial differences between Hume's system and Kant's—and one of the reasons why Kant cannot be an relationalist about concepts—is the different work that each assigns to concepts.

Before moving on to our examination of Kant's theory of concepts, though, we have one more piece of business to take care of. Both Modern and contemporary forms of relationalism about conceptual content must at some point confront the problem of the

⁵³Of course, this object may—and in Hume always is—a further mental entity, but this will not be *essentially* so. What is essential to the content of an idea is not its relation to any other idea, but its relation to its *object*.

unity of the proposition.⁵⁴ As it is, in part, a dissatisfaction with Modern attempts to solve this problem that motivates Kant's inferentialism, and as one of the arguments in favor of that inferentialism is that it seems to be able to solve this problem, it will be worthwhile to spend a small amount of time seeing how this problem arises for Hume in light of his theory of abstract ideas.

The problem of the unity of the proposition is simply the challenge to say what differentiates a proposition (or a belief, a judgment, a sentence, etc.) from a list of names. The problem is particularly acute for the relationalist about conceptual content because it is an immediate consequence of relationalism that (at least) the concepts that appear in a proposition—and in most cases, the other items as well—each have the content that they do independently of any of the other items appearing in that proposition, and independently of the proposition itself.⁵⁵ For the relationalist, the concepts that appear in a proposition have the content they do by standing in certain relations *to their objects*, not to any other mental/linguistic item. For Hume, the concrete simple ideas that appear in propositions are *copies of their objects*. Complex concrete ideas purport to be copies of their objects. Abstract ideas have as their content the sum of the content of the ideas that they resemble and therefore dispose one to recall upon being prompted.

This being the case, the elements of a proposition each look to be a kind of name, either of an object, a set of objects, etc., and so the problem of how to differentiate a

⁵⁴For treatments of the problem of the unity of the proposition as it appears in Modern philosophy, cf. Linsky, (1992), Ott, (2004). For a treatment of this problem's recurrence at the start of the 20th century, cf. Gibson, (2004). For contemporary treatments of the problem, cf. Carruthers, (1983). Wiggins, (1984), Brandom, (1994), (Fodor, 2003).

⁵⁵Cf. Sellars, (1963b) and Landy, (2007).

proposition from a list of names is particularly pressing.⁵⁶ To see why, consider the following list of names.

(1) Joan, Judy, Jessica, Jeffrey

This list is clearly not a proposition. It does not *say* anything. It does not represent anything as being the case, or any object as being a certain way. If, however, this list is not a proposition, neither are the following.

(2) ‘This sphere’, ‘is red’

(3) An idea of a sphere, an idea of something red

(4) An idea of a red sphere

For a relationalist, ‘this sphere’ and ‘is red’ both have the content that they do by standing in a certain relation to their objects. Simply putting each of these next to one another does not make them into a proposition, and more specifically does not make them into the proposition that this sphere is red. Similarly, on Hume’s account, according to which it is *ideas* that have content by standing in certain relations to their objects, placing two independently contentful ideas next to one another, or associating two such ideas, also does not make a proposition out of them. (2) and (3), like (1), do not *say* anything, do not represent anything as being the case, etc.⁵⁷

It is perhaps because he sees this that Hume himself takes a different line on the unity of the proposition altogether. Hume’s official position is that a proposition—or a belief—is not a unity of distinct items at all, but rather it is a single idea with a great

⁵⁶Hume tries to get by with objects and sets of objects; Plato includes the Forms; Frege includes “unsaturated” concepts, etc.

⁵⁷It is worth noting that it is unclear whether the standard contemporary move—made in, for instance, Fodor, (2003)—of making each of the items on what would otherwise be a mere list of names *syntactically* structured helps at all. It should be obvious that we could also put such items on a list, and thus the question of how to differentiate such a list from a proposition still remains.

degree of force and vivacity.⁵⁸ The role that force and vivacity play here is simply to distinguish a belief from a mere contemplation. What is important to notice for our purposes is that *what* is believed or contemplated is *a single idea*. Hume takes it that all belief (or judgment) concerns

the existence of objects or of their qualities. 'Tis also evident, that the idea of existence is nothing different from the idea of any object, and that when after the simple conception of any thing we wou'd conceive it as existent, we in reality make no addition to or alteration on our first idea. (T 1.3.7.2; SBN 94)

Thus, Hume embraces (4) as best representing the form a proposition. Of course, it should be obvious that this will not do at all. Reducing a list to a single item surely cannot be the way of explaining how it is that a proposition differs from a list, and adding that this single item appear particularly forcefully and vivaciously to us does not help either.

All of this is not to say that Hume does not have additional resources to marshal in addressing the problem of the unity of the proposition, or that the problem is unsolvable for any relationalist.⁵⁹ Rather it is simply to point out that this problem is a standing challenge to *any* philosopher who undertakes to account for the content of concepts, and that there is a straightforward way that it is particularly acute for the relationalist. As we will soon see, this problem is a going concern of Kant's, and is one to which he thinks he has a solution.

While Hume's theory of concepts is fairly straightforward and simple, Kant's is

⁵⁸Hume argues as follows. All belief concerns "the existence of objects or of their qualities"(T 1.3.7.2; SBN 94). "'Tis also evident, that the idea of existence is nothing different from the idea of any object" (T 1.3.7.2; SBN 94). "When you wou'd in any way vary the idea of a particular object, you can only encrease or diminish its force and vivacity"(T 1.3.7.5; SBN 96). "An opinion, therefore, or belief may be most accurately defin'd a lively idea related to or associated with a present impression"(T 1.3.7.5; SBN 96)

⁵⁹Cf. Landy, (2007) for a catalogue of possible Humean reactions to this challenge.

not. In fact, to appreciate the full force of Kant's argument against Hume's relationalism we must plumb at least some of the depths of the notoriously murky Transcendental Deduction of the Pure Concepts of the Understanding, one of the places in his corpus in which Kant pays significant attention to the nature and function of concepts. I will begin this expedition with a brief overview of the purposes and structure of the Deduction.⁶⁰ I will then move on to discuss the role that Kant assigns concepts in our mental lives as a result of the conclusions of the Deduction. Lastly, I will draw out the consequences of this assignment for the inferentialism/relationalism debate.

As is well known, Kant's Transcendental Deduction is his attempt to answer a certain *quid juris* question.⁶¹ As Kant tells us,

Jurists, when they speak of entitlements and claims, distinguish in a legal matter between questions about what is lawful (*quid juris*) and that which concerns the fact (*quid facti*), and since they demand proof of both, they call the first, that which is to establish the entitlement of the legal claim, the deduction. (A84/B117)⁶²

Kant's Deduction concerns a kind of entitlement, and more particularly an entitlement that creatures like us have to the use of a particular kind of *concept*: pure *a priori* concepts. An empirical *a posteriori* concept, for Kant, is a concept the content of which is derived from experience—in a sense with which we will not concern ourselves just yet—and the justification of the use of which is conducted via an appeal to this pedigree.⁶³

⁶⁰In what follows, I will present, but not defend, a particular reading of the Transcendental Deduction that owes much to Sellars, originally, and Rosenberg following him. It is drawing the specific link between this reading of the Deduction and Kant's inferentialism that I take to be the major work of this paper.

⁶¹It is thanks to Henrich, (1989) that we know the full extent to which Kant modeled his deduction on the juridical practices of 18th-century German courts.

⁶²All passages from Kant's *Critique of Pure Reason* are taken from Kant, (CPR).

⁶³Longuenesse, (1998) provides an admirably thorough exploration of this kind of derivation. Rosenberg, (2005) also contains many insights.

We make use of a multitude of empirical concepts without objection from anyone, and take ourselves to be justified in granting them a sense and a supposed signification even without any deduction, because we always have experience ready at hand to prove their objective validity. (A85/B117)

Empirical concepts are, more or less, the concepts with which Hume is most comfortable. They are concepts whose origins can be traced up to experience (although perhaps not in the straightforward way that Hume thinks they can).

Pure *a priori* concepts, on the other hand, are those concepts that are not derived from experience, and the justification of the use of which cannot, therefore, be *a posteriori*. It is Kant's goal in the Transcendental Deduction (and in the *Critique* more generally) to provide an *a priori* justification of such concepts.

Among the many concepts, however, that constitute the very mixed fabric of human cognition, there are some that are also destined for pure use *a priori* (completely independently of all experience), and these always require a deduction of their entitlement, since proofs from experience are not sufficient for the lawfulness of such a use, and yet one must know how these concepts can be related to objects that they do not derive from any experience. (A85/B117)

Pure *a priori* concepts are those concepts, more or less, with which Hume is not comfortable. They are the concepts—e.g., necessity, persistence, etc.—that Hume argues that we *cannot possibly have* because they do not resemble any of our impressions. Hume thus has to explain how it is that we *think* we have such concepts, and what in our mental lives we are mistaking for them. Kant, on the other hand, believes that we do make (legitimate) use of such concepts, and the goal of the Transcendental Deduction is to justify this use.⁶⁴ Since the use to which such concepts—and all concepts—are put is to be applied to objects, what must be shown in this deduction is that applying pure *a priori*

⁶⁴It is not an accident that I switched from the locution 'having concepts' to 'making use of concepts' when I switched from talking about Hume to talking about Kant. Hume, as I pointed out, thinks of concepts as mental entities. Kant, on the other hand, as we will see, thinks of them as *rules*.

concepts to objects is a justified practice.

The way that Kant sets out to show this is by showing *this* practice is an essential part of *another* practice, which is itself justified.⁶⁵ Suppose, for instance, that I have been given permission to play baseball today. Now suppose that the question arises whether I have permission to take an at bat today. By showing that taking an at bat is an essential part of playing baseball, I thereby show that I have permission to take an at bat. That is, one could not have permission to play baseball without *also* having permission to take an at bat.⁶⁶ Similarly, Kant's plan in the Deduction is to show that the use of pure *a priori* concepts is an essential part of another practice that is itself justified, and thereby to secure justification for the former practice.⁶⁷

The practice of which Kant takes the employment of pure *a priori* concepts to be an essential part is just the practice of conceiving of some representations as belonging to *oneself*. That is, Kant argues throughout the course of the *Critique* that employing pure *a priori* concepts is necessary for conceiving of one's thoughts *as* one's own. Here is Kant, early on in the B-Deduction, explicitly stating this as the condition to which he will appeal during the course of the Deduction.

For the manifold representations that are given in a certain intuition would not all together be my representations if they did not all together belong to a self-consciousness; i.e., as my representations (even if I am not conscious of them as such) they must yet necessarily be in accord with the condition under which alone they can stand together in a universal self-consciousness, because otherwise they would not throughout belong to me. From this original combination much may be inferred. (B131)

⁶⁵For a similar approach to a slightly different justificatory problem cf. Sellars (1964) and Sellars, (1988). For a similar approach to the Transcendental Deduction cf. Rosenberg, (2005). For a discussion of Rosenberg's approach cf. Landy, (2009).

⁶⁶Designated hitters, bench warmers, and rain-outs aside.

⁶⁷Technically, this project is not completed until the end of the Analytic of Principles.

The mental representations that are given in an intuition must all be *someone's* mental representations. Kant is concerned with what conditions must be in place for the possibility of a person's being justified in claiming his representations as his own. As the cryptic remark at the end of this quotation implies, it is from the conditions of this practice—the practice of being able to claim one's representations as one's own—that Kant hopes to justify the practice of employing pure *a priori* concepts.

Part of Kant's insight here is to see that being able to claim one's representations as one's own is not as straightforward a process as some of his Modern predecessors thought it to be. Descartes, for instance, seems to think that the inference from a premise of the form,

(D1) [I think x] and [I think y] and [I think z]

leads validly to a conclusion of the form,

(D2) [The I that thinks x] = [the I that thinks y] = [the I that thinks z].

That is, Descartes takes the fact that he can introspectively observe that he thinks x, and that he can introspectively observe that he thinks y, and that he can introspectively observe that he thinks z, to imply that it is one and the same thing, *he*, that has all of these thoughts.⁶⁸ Descartes famously writes of his arrival at this conclusion that,

This is a considerable list, if everything on it belongs to me. But does it? Is it not one and the same 'I' who is now doubting almost everything, who nonetheless understands some things, denies everything else, desires to know more, is unwilling to be deceived, imagines many things even involuntarily, and is aware of things just as true as the fact that I exist, even if I am asleep all the time, and even if he who created me is doing all he can to deceive me? [...] That fact that it is I who am doubting and understanding and willing is so evident that I see no way of making it any clearer. (CSM; 19)

⁶⁸Of course, it is controversial whether *Descartes* actually makes this inference. It will suffice for present purposes to see that it is an inference with which Kant's Modern predecessors are concerned, and to which Kant pays a good deal of attention.

Of course, those of us who have read our Hume find the matter to be significantly less clear. We know that this inference—from the existence of certain experiences to the identity of the subjects of these experiences—is fallacious. Putting the matter first-personally, as Descartes does, Hume writes,

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. [...] The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no *simplicity* in it at one time, nor *identity* in different; whatever natural propension we may have to imagine that simplicity and identity. (T1.4.6.3-4; SBN 252)

What Hume points out here is that, when we introspect, we find exactly the matter that Descartes does—this or that perception—but that this is not sufficient to yield an experience of the *self*—something which endures through time and is the *subject* of these perceptions. As Kant puts it, ‘[the] identity of the subject, of which I can be conscious in all my representations, does not concern any intuition of the subject, whereby it is given as object’(B408). If, however, we can have the experiences that would justify our endorsing Descartes’ premise, but still lack the resources for supporting his conclusion, then clearly his inference is fallacious. Because Hume thinks that an *experience* of the self is the only ground that could warrant the further premise needed to make the argument valid, when he fails to find such an experience, he famously rejects Descartes’ conclusion. Kant, as we are about to see, does not.⁶⁹

⁶⁹Of course, Hume does not *really* reject Descartes’ conclusion either. Famously, in the Appendix to the *Treatise* he writes, “But upon a more strict review of the section concerning personal identity, 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” (T App. 10; SBN 633). The problem Hume finds with his account of personal identity is that if it is to work, there must be some principle according to which the various

What Kant sees is that, although Descartes' inference is invalid, his conclusion is one that each of us is nonetheless entirely justified in accepting. That is, we are each justified, according to Kant, in conceiving ourselves as single, unified subjects of experience persisting through time. This is just what it is to be able to claim various temporally dispersed representations as our own. It is to identify those representations as belonging to a single, unified self persisting through time. Kant takes the claim expressing this identification—that it must be possible for *me* to think all *my* representations collectively as *mine*, the principle of the necessary unity of apperception—to be analytic, and so takes our justification for holding it to be straightforward.⁷⁰ His question is not *whether* we are justified in so thinking of ourselves, but rather *how* we come to be so justified. The lesson that Kant learns from Hume, *contra* Descartes, then is not that we ought not to (or cannot) conceive of ourselves as single, unified subjects persisting through time, but rather that our doing so cannot consist in an experience either of this persisting self or of the manifold of experiences that this subject has. It cannot consist in the former because we have no such experience. It cannot consist in the latter because such a manifold is not sufficient for constituting the idea of a single self that is the subject of the entirety of such a manifold.

Here is Kant expressing both the problem that he finds in Descartes, along with the general outlines for his solution to it.

perceptions that constitute *a self* are united. Because the only principles that Hume is willing to consider are those that can be *experienced*, he is left in despair. “But 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” (T App. 20; SBN 636).

⁷⁰The principle for uniting the self that Hume sought, Kant thinks, is simply ‘*These thoughts are mine*’. There is no *reductive* principle according to which we can identify all and only *my* thoughts. We can only so identify them by *presupposing* the self that has such thoughts. The thoughts that are mine are simply the ones that *I* have.

For the empirical consciousness that accompanies different representations is by itself dispersed and without relation to the identity of the subject. The latter relation therefore does not yet come about by my accompanying each representation with consciousness, but rather by my adding one representation to the other and being conscious of their synthesis. Therefore it is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself (B134)

Kant's explanation, then, of how it is possible to conceive of oneself as a single, unified subject of experience persisting through time is that this is possible only if one can 'combine a manifold of given representations in one consciousness.' What Kant sees is that while (D2) does not follow from (D1), it does follow from,

(K) I think $[x + y + z]$.⁷¹

That is, while introspectively observing a manifold of various representations is not sufficient for conceiving of oneself as a single, unified subject persisting through time, conceiving of a single representation, *the content of which includes a manifold of representations*, is sufficient. Otherwise put, he sees that we would be justified in inferring that one and the same thing experiences all of x, y and z if we were justified in thinking that one and the same thing thinks something *else* whose *elements* included x, y and z. Again, if x, y and z were three parts of *a single cognition* had by a single individual, then it would follow trivially that the I that thinks x is the same as the I that thinks y and the same as the I that thinks z.

As we should now expect, the cognition that Kant thinks plays this role in our thought is exactly the kind of cognition that necessarily employs pure *a priori* concepts. Thus, Kant's strategy can now be fleshed out a bit more. Kant sets out to justify our use

⁷¹The nature of the '+'s here will be our topic later. For now these just signal some sort of combination of a manifold of intuitions occurring as part of a single cognition. Later we will see that this combination is *inferential*.

of pure *a priori* concepts. His plan is to show that our use of pure *a priori* concepts is an essential part of our engaging in another practice that is itself justified. This practice is that of conceiving of ourselves as single, unified subjects of experience persisting through time. Following Hume, he argues that doing *this* is not possible via an experience of such a self, because we have no such experience. He further follows Hume in thinking that being introspectively aware of each member a manifold of experiences is also not sufficient for these purposes. Most recently we have seen Kant notice that it *would* suffice for so conceiving ourselves to have a single cognition, the contents of which are the set of experiences in need of uniting. If it is true that we can only have such a cognition *by* employing the pure *a priori* concepts, and the rest of Kant's arguments here are sound, then he will have found the justificatory argument for which he is searching.

Our next task will be to see what reason Kant gives for thinking that the pure *a priori* concepts are necessary for having this kind of cognition. Before we can do that, however, we must see what kind of cognition Kant thinks this to be. He writes,

Understanding is, generally speaking, the faculty of cognitions. These consist in the determinate relation of given representations to an object. An object, however, is that in the concept of which the manifold is united.
(B137)

Kant here suggests that our manifold of intuitions is united by our employment of *object-concepts*.⁷² The single cognition, $[x + y + z]$ is the use of an object-concept whose elements include a manifold of intuitions and which, because it is a single cognition, is had by a single thinker.⁷³ An example will help illustrate why it is that Kant thinks that

⁷²I here take over Kant's terminology and speak of a manifold of *intuitions*, rather a manifold of experiences.

⁷³It is, more specifically, the application of an object concept to a manifold of intuitions. Such an application occurs in a *judgment*, and explaining how *this* works, how concepts are "applied" to manifolds

this particular kind of cognition can do this work, while other kinds cannot. Suppose that one is presented with the following diachronic manifold of intuitions.

- t₁: This short grey tail.
- t₂: This big grey body.
- t₃: This big, flat grey ear.
- t₄: This long grey trunk.⁷⁴

According to Kant, it is by thinking of such a manifold of intuitions using an object-concept, such as ‘elephant’, that we unite them in a single cognition.⁷⁵ Our question is what makes such a concept particularly suited to do this work. Kant’s answer is that by applying an object-concept to a manifold of intuitions, by thinking of these intuitions as being of an *elephant*, what we crucially add to that manifold is an element of *necessity*.

If we enquire what new character *relation to an object* confers upon our representations, what dignity they thereby acquire, we find that it results only in subjecting the representations to a rule, and so in necessitating us to connect them in some specific manner; and conversely, that only in so far as our representations are necessitated in a certain order as regards their time-relations do they acquire objective meaning. (A197/B242)

By applying the object-concept ‘elephant’ to the above manifold of intuitions, by conceiving of that manifold as being the result of an encounter with an elephant, we

of intuitions, will bring us full circle back to where we started our investigation of Kant: the problem of the unity of the proposition. We have a way to go yet, however, before we complete the circle.

⁷⁴Notice that we could have run this example synchronically, if what we wanted was a unified self at a time. For instance, we could have made our manifold out of the synchronic experiences (1) I see a grey trunk, (2) I hear a loud trumpeting sound, (3) I feel leathery skin, (4) I smell dung, etc., or even synchronic experiences such as (1) I see a grey patch in such-and-such a portion of my visual field, (2) I see a darker grey patch in such-and-such other part of my visual field, etc.

⁷⁵It is worth noting here that the terms in which we have described these intuitions also make use of object-concepts—‘tail’, ‘body’, ‘ear’ and ‘trunk’. This gets us into the complicated theme of the productive imagination and the difference between intuitions and *sensations*. Intuitions, for Kant, are already themselves *conceptual* episodes—which for us means that they already enmeshed in a network of inferential relations. Sensations are the non-conceptual episodes that are united by the understanding to form intuitions (cf. A79/B104), and so it is sensations that is, strictly speaking, the closer analogue to Hume’s impressions. Hanna, (2001) is a recent attempt to push intuitions further towards the non-conceptual.

suppose that those intuitions are connected to one another *necessarily*.⁷⁶ We suppose that it is not an accident that we encounter first a tail, then a body, then an ear, then a trunk.⁷⁷ We attribute the order of this series of intuitions to the two-part story of *my* running my gaze from the back to the front *of an elephant*. Kant's thought is that we conceive of ourselves as single, unified subjects of experience persisting through time *by* conceiving of our manifold of intuitions as being the result of multiple encounters with a single, lawful world of objects persisting in space and time.

Just this transcendental unity of apperception, however, makes out of all possible appearances that can ever come together in one experience a connection of all of these representations in accordance with laws. For this unity of consciousness would be impossible if in the cognition of the manifold the mind could not become conscious of the identity of the function by means of which this manifold is synthetically combined into one cognition. (A108)

Thus the single cognition that unites all and only those actual and possible representations that are properly called *mine* is just the thought of this single world as encountered *by me*. Since this thought unites the entire manifold of intuitions, and since this single cognition is had, necessarily, by a single self, the possibility of having such a cognition allows one to conceive of oneself as a single, unified self persisting through time. Object-concepts are just those concepts through which we form such a thought piecemeal, as we encounter not the entire world, all at once, but the objects that make it

⁷⁶It is not, of course, the case that an elephant *necessarily has* a tail, an ear, a trunk arranged in such-and-such a configuration. There are, sadly, elephants without such parts. Rather the necessity attaches to the relation of these *intuitions* to one another. They are necessarily connected, as opposed to being merely contingently *associated*, because they are the result of an encounter with an object, in this case an elephant.

⁷⁷Kant argues, most explicitly in the *Analytic of Principles*, that the difference between conceiving of a manifold of intuitions as a mere series of random representations, and conceiving of such a manifold as a series of representations *of objects*, is thinking of the items of that manifold as being necessarily connected to one another. Cf. Strawson, (1959): chapter 2.

up, one at a time.⁷⁸

We can now see how Kant's strategy in the Deduction is carried out, if not in detail, then at least in broad strokes. The pure *a priori* concepts are the rules that specify what counts as an *object*-concept, that is, what counts as a concept capable of positing the sort of necessity among intuitions that results in the proper uniting of that manifold.⁷⁹ Thus, the employment of empirical concepts in accordance with the pure *a priori* concepts is necessary for conceiving of a single, lawful world of objects persisting in space and time. This, in turn, is necessary for conceiving of ourselves as single, unified subjects of experience persisting through time. Conceiving ourselves in this way is a practice that is analytically justified. And so, the employment of pure *a priori* concepts is justified on the grounds that it is an essential part of a practice that is itself analytically justified.

What should interest us most in this picture is the role that Kant here assigns to object-concepts. Their purpose is twofold. First, insofar as we apply an object-concept to a manifold of intuitions, we take those intuitions to be necessarily connected to one another. An object-concept is that which, somehow, posits these necessary connections. Second, in deploying a particular object-concept in this way, we also deploy a certain explanation of *why* the necessary connections amongst these intuitions exist; it is because they are the result of an encounter with an object that is part of world governed by such necessities. Thus, for Kant, the primary work that concepts do in our mental lives is to unite our manifold of intuitions.

⁷⁸There is, of course, a long story to tell here about the *ways* in which we encounter this world of objects: from a perspective in space and time.

⁷⁹Cf. Sellars, (2002).

The most important point about Kant's theory of concepts that we can take away from the Transcendental Deduction, then, concerns the role of concepts vis-à-vis the positing of necessities amongst the elements of the manifold of intuitions. Suppose Kant is correct to think that concepts must serve this function. What we need, then, is an account of *how* they do so. How is it, that is, that concepts "posit" these necessities? How do they represent the elements of the manifold as being necessarily connected in these ways? Here is Kant's answer.

A concept is always, as regards its form, something universal which serves as a rule. The concept of body, for instance, as the unity of the manifold which is thought through it, serves as a rule, in our cognition of outer appearances. But it can be a rule for intuitions only insofar as it represents in any given appearances the necessary reproduction of their manifold, and thereby the synthetic unity in our consciousness of them. (A106)
(A106)

Our concept of an object serves as a *rule* in our cognition of outer appearances, in our thought about objects. As what kind of rule?⁸⁰ What I want to argue for presently—on both exegetical as well as philosophical grounds—is that Kant is thinking of rules here as rules of *inference*, and that his theory of concepts is an inferentialist one.⁸¹ The picture, when it is completed, will be this. We represent necessities among the elements of the manifold of intuitions at the object-level by licensing certain material inferences at the meta-level. To borrow an example from Wilfrid Sellars, it is by licensing the inference from 'lightning now' to 'thunder soon' that we come to represent the necessary connection between lightning and thunder. It is because our inferential practices are such as to treat certain material inferential moves as valid, that we come to represent the world

⁸⁰It is surprising how infrequently this question is asked in the literature on the *Critique*, the Transcendental Deduction, and Kant's theory of concepts more generally. Notable exceptions are Sellars, (1967), Pippin, (1982), Longuenesse, (1998), and Rosenberg (2005).

⁸¹That is, that not only is a concept an inferential rule, for Kant, but that is *all* that it is.

as one of objects bearing certain necessary connections to one another.

I will begin with the exegetical part of the argument.⁸² I should say at the outset that I will not argue that the following texts *prove* that Kant is an inferentialist. This is, admittedly, not the only possible reading of which these passages admit. Rather, I will argue that these passages are suggestive of a more comprehensive theory of concepts than is explicitly introduced in them, and that the best sense that can be made of what this theory might be is that it is a (very specific kind of) inferentialism. In particular, it is an inferentialism that, combined with Kant's sophisticated account of perception, yields a kind of picture-theory according to which intuitions are the pictorial elements, and concepts are the relations into which these elements are put in order to represent the relations in which the items pictured stand. More on these details later. For now, let's turn to the text.

The place in the first *Critique* where Kant is most explicit about his theory of concepts is in the Metaphysical Deduction, which is fairly short, and will repay a close reading. The first passage of interest is the following.

Now the understanding can make no other use of these concepts than that of judging by means of them. Since no representation pertains to the object immediately except intuition alone, a concept is thus never immediately related to an object, but is always related to some other

⁸²I will *not* be looking at all at the *Jäsche Logic*, which many commentators argue presents compelling evidence that Kant held an abstractionist theory of concepts and concept acquisition. There are two main reasons for this omission, which I will present, but for which I will not argue here. The first is that much of what is uncontroversial in what Kant has to say about his theory of concepts in the first *Critique* is fairly clearly incompatible with the abstractionist theory of concepts presented in the *Logic*. For instance, the anti-empiricist doctrine of the *spontaneity* of the understanding pretty clearly rules out the possibility that concepts are products of mechanical operations performed on intuitions (or anything else). Secondly, given this incompatibility, we have more reason to take seriously what Kant has to say about concepts in the first *Critique* than we do what he says in the *Logic*. As has been often noted, the *Logic* is a compilation (by Gottlob Benjamin Jäsche) of the notes that Kant used in his exoteric lectures on logic at the university. It is, therefore, true both that he did not have the final say on what went into the published version of these notes, and that he had reason to present a simplified and unrepresentative version of his *very* complex views on concepts in these lectures. For these reasons, I will proceed by taking the first *Critique* as the final word on Kant's theory of concepts.

representation of it (whether that be an intuition or itself already a concept). Judgment is therefore the mediate cognition of an object, hence the representation of a representation of it. (A68/B93)

The first thing we are told here is that the understanding can make no other use of a concept than to judge by means of it. It should be clear that the inferentialist can, at least, make good sense of this claim. If concepts are given their meaning via the role they play in a system of inference, then a concept outside of a judgment—where it cannot license, prohibit, or require such inferences—has no use. Furthermore, the specific use that Kant attributes to concepts—judging by means of them—also makes perfect sense on an inferentialist line. It is because concepts license, prohibit, and require certain judgments that, when one judges, one judges by means of these concepts. It is in accord with the rules that such concepts provide that one ought to make one's judgments.

The passage goes on to declare that an intuition is a representation of an object, and that a judgment is a representation of an intuition. It seems clear that the judgment here is a representation of an intuition *because* it has a concept in it. That is, it is the presence of a concept in a judgment that *makes the judgment* a representation of an intuition. In what sense, then, we must wonder, does a judgment represent an intuition, and how does the concept deployed in a judgment accomplish this. What follows this passage seems to be Kant's explanation of just this.

So in the judgment, e.g., "All bodies are divisible," the concept of the divisible is related to various other concepts; among these, however, it is here particularly related to the concept of body, and this in turn is related to certain appearances that come before us. (A68/B93)

Note that Kant writes that *in the judgment*, "All bodies are divisible," the concept of the divisible is related to *various* other concepts, and that the concept of body is only one among these. This is interesting because on its face this judgment relates only *two*

concepts: bodies and divisibility. Kant, however, writes that even in this very judgment, despite these appearances, 'divisible' relates not only to 'bodies', but also to some other, unnamed concepts. What I want to suggest is that this is can best be understood by supposing that something about *the concept 'divisible' itself* relates it to other concepts. More particularly, what I want to suggest is that the concept 'divisible' just *is* its relation to other concepts.⁸³ The work that 'divisible' does in the judgment, 'All bodies are divisible,' is exhausted by the relations that it bears to other concepts (and to those certain appearances that come before us). To place an 'are divisible' next to an 'all bodies' is to take up a certain position in a network of various related concepts and intuitions (a network of various *normative* relations: commitments, prohibitions, permissions, etc.). It is to judge that bodies are related in a certain way to not only divisibility, but all that being divisible entails, e.g., being extended, having parts, etc.

The other important part of this passage to note is the *other* direction of flow of concepts implied in it. That is, not only is 'divisible' connected to other concepts, but equally importantly, 'body' is connected to "certain appearances that come before us". What Kant is implicitly up to here is presenting for the first time his *two-part* inferentialism. The first part of this theory is that a concept has its content by being related (inferentially) to other concepts. The second part is that this network of inferentially related judgments must eventually include "certain appearances that come

⁸³This interpretation is thus crucially different from that presented in Hanna, (2001). Hanna argues that while concepts *also* act as rules of inference this is by virtue of their representing intrinsic structural *properties* of empirical objects. There is not space enough to show it here, but I would argue that such a reading makes Hanna's Kant a kind of relationalist, and therefore susceptible to the problem of the unity of the proposition.

before us”.⁸⁴ That is, the operation of concepts in our mental lives presupposes that these concepts are connected not only to each other, but also to *intuitions*, which provide their essential link to perception, and thereby to the objects which both of these kinds of representations, in their own way, represent.⁸⁵

The concept of body thus signifies something, e.g., metal, which can be cognized through that concept. It is therefore a concept only because other representations are contained under it by means of which it can be related to objects. (A69/B94)

Not only are the relations that concepts bear to one another a crucial part of Kant’s story, but the relations that they bear to intuitions is equally important. This is because without this link to intuitions, and thereby to perception, concepts would have no connection to objects, which as we are about to see, and as Kant suggests here, would rob them of their representative power. Intuitions provide the necessary starting place of cognition, or conceptual thought. Kant has essentially discovered language-language and language-entry moves, and made them part of the very content of a concept.⁸⁶

What I am suggesting, then, is that the best way to understand Kant’s claim that judgments are meta-representations is via the additional thesis that the way that we represent, e.g., a body as being divisible is by placing an intuition of a body into a system of inference that licenses certain further judgments about bodies: e.g., all bodies are

⁸⁴It is by including these certain appearances that come before use that Kant signals that he is here considering *empirical* concepts. *Pure* concepts do not have, or need to have, this empirical component. Pure *a priori* concepts, the Categories, are a different beast still. They can plausibly be read, I think, as inferential meta-rules concerning what first-order rules count as rules for understanding manifolds of intuitions *as objects*.

⁸⁵There is not space to argue it here, but I will point out that if concepts are related *inferentially* to intuitions, which I think they are, then intuitions must, in some sense, be conceptual. Room, however, is still made for the non-conceptual in our mental lives by distinguishing intuitions from *sensations*. Cf. Sellars, (1967).

⁸⁶Of course, Kant is not thinking in terms of *languages*, but rather systems of mental representation. For the story of how philosophy moved from this Kantian perspective to one that explicitly engages language and communal standards for the correctness and incorrectness of inference cf. Landy, (2008).

extended. Of course, what Kant *says* is just that in the judgment, “All bodies are divisible,” the concept ‘divisible’ is connected to various other concepts, and that the concept body is connected to various appearances (intuitions). The key to my exegetical thesis is that the best way to understand *this* as an *explanation* of what comes before it—the claim that judgments are meta-representations—is via the two-part inferentialism I have been outlining.

Here is another way of approaching these passages. Kant claims that a judgment is a meta-representation: a representation of an intuition. Now, clearly not all judgments are *about* intuitions. (While we can make judgments about intuitions, the more paradigmatic case is one in which the judgment is about objects in space and time.) So, typically, it is *by* representing intuitions in a certain way that a judgment comes to represent *the world* as being a certain way. The key questions here are how do judgments represent intuitions, and how does doing this represent the world as being a certain way? The inferentialist has a ready answer. Necessary connections among worldly objects are represented by placing the elements of the manifold of intuitions into certain relations: inferential ones (that is, not relations with their objects, but with other mental/linguistic items). These relations mimic, or picture, the relations that the objects pictured stand in to each other. Intuitions are the representative counterparts of objects, and concepts ways of placing intuitions into relations with each other that are the counterparts of relations in which their objects stand. Consider what Kant says in another context, again about the concept body in slightly different context.

Thus in the case of the perception of something outside of us the concept of body makes necessary the representation of extension, and with it that of impenetrability, of shape, etc. (A106)

Obviously, when Kant writes here that ‘the concept of body *makes necessary* the representation of extension’ he does not mean that when we think ‘body’ we necessarily also think ‘extension’. Clearly, there are plenty of times in the life of a human being that one thinks (and *can* think) ‘body’ without also thinking ‘extension’. Kant’s claim is not a claim about the *association* of ideas (a la Hume), but rather can best be understood as a claim about the inferential, *normative* relations between concepts. This is why Kant casts concepts as *rules*. When one judges of something that it is a body, one is also thereby *committed to* judging of that thing that it is extended.⁸⁷ This is a key difference between Kant’s inferentialism and Hume’s associationism. Kant leaves open the possibility that while we might, in fact, judge falsely that some body is not extended, we are necessarily wrong to do so.

Concepts, on this line, are the rules that govern such inferential connections. For instance, our concept of an elephant, on this view, is just a rule for putting a manifold of intuitions like the one we encountered earlier into the right relations with each other, and with a set of other possible intuitions of elephants and other elements in the natural world. It is by doing *this* that we use the concept of an elephant to *unite* all of these intuitions, and it is in this sense that in any *particular* judgment that employs the concept elephant that concept is already, *in that very judgment*, related to other concepts (as well as certain intuitions, and other judgments).

Before we move on, it is important to note here (as earlier) that while some of the examples we have been lately using are examples of inferences that are plausibly

⁸⁷A similar reading is available of Kant’s talk of one concept’s *containing* another. A concept, *A*, can be said to contain another, *B*, just in case if one commits oneself to something’s being *A* one also thereby commits oneself to its being *B*. This, of course, would need to be worked out in more detail alongside Kant’s texts, seems like a *prime facie* good start.

considered *analytic*. Kant—and the inferentialist more generally—is not limited to including only these among the content-determining inferences for a given concept. ‘Body’ is necessarily connected to ‘extension’ and thereby to ‘shape’, ‘impenetrability’, etc. Such inferences may well be analytic. Consider, however, the connections between the intuitions of a short grey tail, a big grey body, and a long grey trunk. We wanted to cast these intuitions as being united by the application of the concept ‘elephant’ to them. What we are in a position to see now is that the way that these intuitions are so unified is *inferentially*. When one encounters a short grey tail and judges of it that it is the short grey tail of an elephant, one is thereby committed to also judging, *ceteris paribus*, that this tail is connected to a big grey body (of which one can, in appropriate circumstances, also form an intuition), and that this big grey body is connected to a head with a long grey trunk, etc. Prime facie, these latter kinds of inferences are not analytic.⁸⁸ That is, they are not instance of *formally valid* inferences, but rather are examples of what Sellars calls *material* inference. It is material inferences that allow us to represent the lawful relations between spatio-temporal objects (like the parts of elephants, or lightning and thunder, etc.). It is by licensing the inference from ‘lightning now’ to ‘thunder soon’ that we represent the necessary connection, the lawful relation, of lightning and thunder. Kant would here add that not only do physical necessities fall under this rubric, but so do those of mathematics and transcendental philosophy.

What all of this amounts to is not just a new theory of how the content of concepts is articulated (inferentially as opposed to relationally), but also a new conception of how it is that a concept functions in our mental lives. Remember on Hume’s account, and on

⁸⁸Sellars, (1963d) and Brandom, (1994) both contain what I take to be plausible ways of cashing out the analytic/synthetic distinction in post-Quinean ways.

those accounts that are most typical of relationalists, a concept has the content that it does in virtue of its standing in the relation that it does to its object. A concept, on that view, *stands for* an object, or collection of objects, etc. just as an idea stands for its object. It is exactly this thesis that Kant here denies, and whose denial constitutes his strongest argument for inferentialism. For Kant, a concept does not *stand for* an object or collection of objects, etc. at all. Rather, the essential role of a concept is to represent objects *as being a certain way*, as standing in certain necessary relations to one another.⁸⁹ It accomplishes this not by *standing for* these relations, but by placing intuitions (which do stand for their objects) in certain relations with one another. Thus Kant has something like a *picture theory* of meaning, according to which the *elements* of the picture are intuitions, and the *relations* into which these elements are put, their form, is constituted by the concepts under which they are subsumed.⁹⁰

Kant draws this distinction in ways of representing as constituting part of the difference between intuitions and concepts.

All intuitions, as sensible, rest on affections, concepts therefore on functions. By a function, however, I understand the unity of the action of ordering different representation under a common one. (A68/B93)

Intuitions are the product of the mind's being affected in certain ways by objects. They are representations of determinate individuals, their objects (which, since Kant is an *empirical realist*, are first and foremost *physical* objects). Concepts, by contrast, are

⁸⁹This explains what Longuenesse calls the "privileging of predication" in Longuenesse, (1998): 104. As Hanna, (2001) points out, this is a substantial and controversial claim that does a great deal of work in Kant's theory of judgment.

⁹⁰Notice that calling the theory described here a picture theory does not thereby commit one to a relationalist account of the representative power of pictures. Clearly, one cannot be an *inferentialist* about such a picture, but one is not thereby forced to relationalism. One might hold, for instance, that a picture represents in virtue of the representative power of its elements, and that each of these represents in virtue of its place in the picture. For a discussion of the relation of inferentialism, picture theory, and relationalism, cf. Rosenberg, (2007): chapter 5.

functions, or rules, for ordering such intuitions. They represent these objects as being connected in certain ways, not by representing some *further* object, but by placing these representations, intuitions, into certain (inferential) relations with one another. (The job of a *function*, in this context, is to carry one from some input—an intuition—to an output of the same kind—other intuitions.) The relationalists' mistake, Kant contends, is that they misunderstand the proper role of concepts in our system of thought. As Kant puts it, concepts do not relate to objects immediately, but only mediately. They do so by relating intuitions to *one another*.

Consider again, then, the judgment, 'This sphere is red'. We saw earlier that the relationalist confronts a difficulty in accounting for the unity of this judgment. Because 'this sphere' and 'is red' have the content that they do in virtue of standing in certain relations to their objects, it is unclear how these *conceptual* representations combine to make the *propositional* representation that this sphere is red. It is unclear, that is, how these representations combine to form more than a mere list of representations.

Kant's answer is that 'is red' does *not* have the content that it does by standing in a certain relation to an object, set of objects, etc. 'This sphere is red' is not, on Kant's view, a combination of two items each of which stands for a certain bit, or certain bits, of the world. Rather, for Kant, a judgment is a *way* of representing *one* item, of saying *how* that item is. It does this *by* representing a representation of it. Placing an 'is red' after an intuition such as 'this sphere' is a way of relating 'this sphere' to other *intuitions*. The items related in this judgment are, in a sense, not an intuition and a concept, but rather multiple intuitions. This is because, as we saw, a concept is a *function*; it takes one from an intuition to various other intuitions (sometimes by way of further concepts/functions).

It does not itself stand for anything, but rather relates items, which themselves stand for objects, to other such items. A concept is a way of signaling certain inferential licenses at the meta-level, and of thereby positing certain necessary connections at the object level.⁹¹ It is *by* relating intuitions inferentially that concepts contribute to a picture of how natural objects are related (necessarily connected). The problem of the unity of the proposition is solved by conceiving concepts as this kind of intra-mental, normative *relation*.

To return to the eponymous topic of this essay, the Transcendental Deduction begins with the necessary goal of conceiving of ourselves as single, unified subjects of experience persisting through time. Kant there argues that this can only be accomplished via employing an object concept in a judgment as a way of uniting a given manifold of intuitions. This unification is accomplished by placing those intuitions into inferential relations with one another. This placement is accomplished by pairing intuitions with concepts, which are themselves nothing but such inferential placeholders. As it turns out, then, the solution to the problem addressed in the Transcendental Deduction is *also* a solution to the problem of the unity of the proposition. This is because the solution to the former casts concepts as achieving the unification of a manifold of intuitions not by standing in any relation to some further object, but rather by placing intuitions into inferential relations with each other. The content of a concept, then, is determined entirely by the inferential role they assign to the intuitions with which they are paired. A judgment cannot be a list of names because concepts *are not names*. Concepts are rules, or functions. The ‘is red’ in, ‘This sphere is red’, does not name a property, redness, but rather signals that the intuition next to it, the ‘this sphere’, is connected to various other

⁹¹Obviously this reading of Kant owes much to Sellars, (1963b), and Sellars, (1979). This is in large part because Sellars is such a close and careful reader of Kant.

intuitions and concepts in specific ways.

A concept is a rule for connecting intuitions, which connections create a picture of the world. The intuitions in this picture represent the objects in the world; the inferential relations between such intuitions represent the necessary connections between these worldly objects. This is exactly what the Transcendental Deduction required for conceiving of oneself as a single, unified subject of experience persisting through time, and it is in this way that the Deduction provides a key justification for what I have argued is Kant's inferentialist account of concepts.

Chapter Four

Sellars on Hume and Kant on Representing Complexes

In his graduate-seminar lectures on Kant—published as *Kant and Pre-Kantian Themes* (Sellars, 2002)—Wilfrid Sellars argues that because Hume cannot distinguish between a vivacious idea and an idea of something vivacious he cannot account for the human ability to represent temporally complex states of affairs. The first section of this paper aims to show that this argument is not properly aimed at the historical Hume who can, on a proper reading, distinguish these kinds of representations. This is not, however, Sellars' only argument for this conclusion. The next section of this paper continues with a discussion of an argument that Sellars presents on Kant's behalf in *Science and Metaphysics*, and its key Kantian premise that, contra Hume, only *conceptual* representations can represent any complex states of affairs as such. The conclusion of this discussion is that Sellars does indeed present compelling reasons for rejecting a Humean account, reasons centered on the *ambiguity* of the associative structure of mental representation (which Sellars subsequently replaces with *inferential* structure). The paper concludes with an examination of the sense in which non-conceptual representations—such as those that Hume considers—represent and the role that these play in conceptual representation for both Sellars and Kant.

In the course of his lectures on Kant's accounts of space and time Sellars asks us to consider how Hume might account for the temporally complex experience of a person raising his hand.⁹² How is it, on Hume's account, that we represent the sequence of events that together comprise a hand being raised? To begin with, we can surely say that this account will involve perceptions (impressions and ideas) of a person with his hand in various *stages* of being raised.

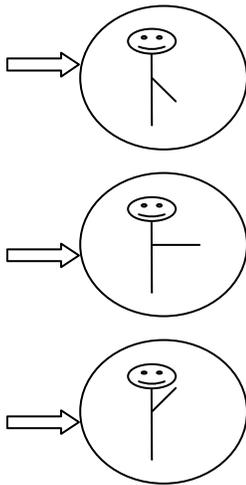


Figure 1

The arrows here indicates that what appears inside the bubbles are what is represented by some perception. So, what we have here are three representations, each of a person with his hand in a different stage of being raised. Of course, none of *these*, Sellars points out, is the representation of *a hand being raised*. The top perception is of a person with his hand down; the middle perception is of a person with his hand parallel to the ground; the last is of a person with his hand raised. None, *by itself*, is a representation of a hand being raised. This much should be obvious.

As Sellars points out, the first step in Hume's account of temporality will be to notice that these representations themselves occur in a temporal sequence.

⁹²We will, for the purposes of this section, be considering representations of temporal complexes—states of affairs with distinct temporal parts—only. In the next section, we will turn to considerations meant to generalize to representations of all kinds of complex states of affairs.

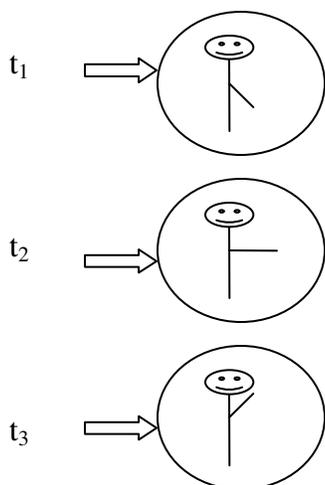


Figure 2

Of course, a temporal sequence of representations is not yet a representation of a temporal sequence. As in the case above, we do not here have a representation of a hand being raised. What we have instead, again, is a representation of a lowered hand, followed by a representation of a hand parallel to the ground, and a representation of a raised hand. Certainly, the fact that these representations occur in a temporal sequence does nothing to add to any of them *individually* the content ‘a hand being raised’; they are each still only representations of hands in various positions (which may, or may not, be part of a hand raising; none of them is a representation of a hand *being raised*. Furthermore, the *collection* of ideas, now themselves explicitly placed in a temporal sequence also does not represent a hand being raised. Hume would certainly agree that there are many sequences of representations that run through the human mind, not every one of which is a representation of a sequence of events. It follows, then, that occurring in such sequence is not sufficient for representing a temporal complex, although it *might*

be necessary for this.⁹³ A sequence of representations, by itself, is not yet a representation of a sequence.

The problem here is that these representations remain *independent* of one another. What we need is not just a representation of each of the stages of a hand-raising, but also a representation that, in some sense, *includes* these stages, and represents them as *parts* of a *whole*, moments in an event of hand-raising. We will not only need a representation of each stage, but also a representation of the relations that they bear to one another. To accomplish this, we will need some account of how these earlier perceptions are *retained* and placed alongside the later ones. We need the earlier representations to be *remembered* later, so that they can be *combined* with the later ones into a complex representation of the complex state of affairs that is a person raising his hand.⁹⁴ As Sellars rightly points out, “Hume speaks in terms of vividness here” (Sellars, 2002: 111). Here is Hume, in the context of distinguishing the memory from the imagination, on the relation of force and vivacity to memory.

The more recent this memory is, the clearer is the idea; and when after a long interval he wou'd return to the contemplation of his object, he always finds its idea to be much decay'd, if not wholly obliterated. We are frequently in doubt concerning the ideas of the memory, as they become weak and feeble; and are at a loss to determine whether any image

⁹³In the next section we will see Sellars and Kant argue that a sequential ordering of perceptions (or intuitions) is *not* necessary for representing temporal complexes, although *some* ordering or structure is. (Actually, both a conceptual and a non-conceptual structure are needed, eventually.)

⁹⁴More precisely, this is a feature that is necessary for the kind of account with which Hume, Kant, and Sellars are concerned, an account in which complex states of affairs are represented by complex representations. (The elements from which these complexes are constructed for each of these philosophers, of course, are quite different. For Hume this basic element is the simple perception; for Kant it is the intuition; for Sellars it is the ‘this-such.’) For Hume and Kant, I suspect that such an account just appeared natural. For Sellars, such an account is a necessary for preserving his nominalist ontology. Cf. Seibt, (1990). There are some contemporary philosophers who propose an alternative to this kind of account. For instance, Fodor argues that many temporally—and generally—complex states of affairs are represented by concepts that are themselves simple, i.e., without parts that are themselves concepts. Cf. Fodor and Lepore, (1992).

proceeds from the fancy or memory, when it is not drawn in such lively colours as distinguish that latter faculty. (T 1.3.5.5; SBN 628)

A memory is an idea, and as such, is a copy of an impression. As time passes, our memories lose more and more of the vivacity that they have at the time of their formation. They become weak and feeble. This is Hume's account of how past impressions are available to the mind for combination with present impressions, and of how the two are distinguished. An impression is retained in the form of an idea; this idea persists, even as the impression is lost, and is available at later times, with a loss of force and vivacity, for combination with other ideas and impressions.

What we need now is Hume's account of what the nature of this *combination* is, of the way the weak and feeble ideas of memory are combined with the forceful and vivacious impressions of the present in order to form a representation of a temporal sequence. Sellars offers us two pictures of this combination. The first looks like this.

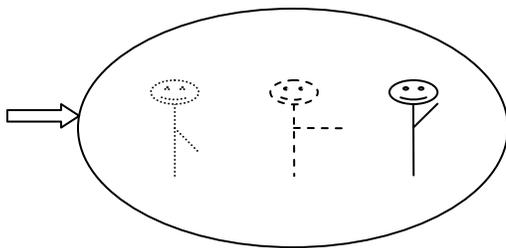


Figure 3 (Sellars, 2002: 111)

Now, the content of an idea that would be pictured like *this* would be a faint *person* with his hand down *standing next to* a slightly less faint person with his hand parallel to the ground standing next to a vivid person with his hand raised. It is clear from the context that Sellars does not mean to accuse Hume of representing temporal sequences by representing temporal parts *as* spatial parts. Rather, what Sellars sets out to picture here is three representations, *in some way combined with one another* into a whole in which all three representations appear. Let us represent this as follows.

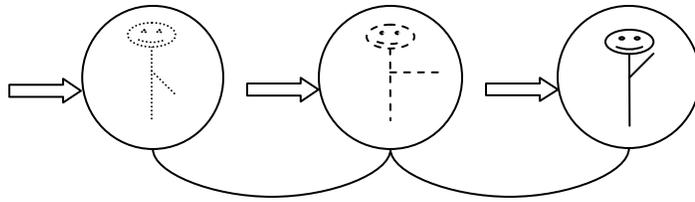


Figure 4

Here we have three representations combined in some way, we do not yet know how. At this point in our story all we know is that it Hume's hope that there is some way of combining representations so as to make the subsequent complex representation a representation of a complex. Still, though, in Figure 4, the ambiguity between a faint idea of a person, and an idea of a faint person remains. As Sellars sees it, this is entirely appropriate.

Consider Sellars' *second* picture of how Hume would attempt to represent a temporal sequence, in which, again, Sellars has Hume combine the constituent representations spatially.

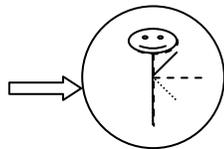


Figure 5 (Sellars, 2002: 112)

What we have here is a complex representation in which the various components *overlay* one another, so that, as Sellars points out, it becomes unclear whether this representation is of the motion of an arm moving upwards, or of a person with three arms each more faded than the next. What is important here, however, is not the confusion concerning the number of arms represented; this can be clarified, at least temporarily, by the convention employed above for picturing unknown relations. Rather what is of interest is the ambiguity that Sellars again attributes to Hume between a faded representation and a representation of something faded. It is because of *this* ambiguity that at this point Sellars abandons Hume's attempt to account for a temporal sequence altogether.

Now this is what posed the problem and, as I said, Hume was really unable to account for it. He really didn't have any account to give. (Sellars, 2002: 113)

Sellars' thought is that Hume simply has no resources here, and that this criticism is clearly decisive. If Hume's account does not distinguish a representation of a faded hand from a faded representation of a hand, then—given that it was the force and vivacity of a perception that was supposed to account for its placement in a representation of a temporal sequence—Hume's account is hopeless. Force and vivacity simply cannot play both the roles that they need to as properties of both *representings* and *representeds*. Their doing so introduces an ambiguity that is fatal to Hume's system.

Of course, this raises the question *for us* of whether Hume really does intend force and vivacity to play both these roles, and whether it is possible for him to avoid this ambiguity. What I will suggest in what follows is that this was never Hume's intention, and there is no reason to think that his account allows for this ambiguity at all.

There are, then, two points in Sellars' construal of Hume that we must keep an eye on in our own discussion of Hume. The first point concerns the structure that complex representations must have to be representations of complex states of affairs as complex states. As of yet, we only have a promissory note issued on Hume's behalf to the effect that there must be *some* structure that does this work. The second point concerns the ambiguity between a forceful and vivacious idea and the idea of something forceful and vivacious. My plan for the remainder of this section is to pick up the latter topic first, as its resolution will provide some insight into the resolution of the former.

The most prominent feature of Sellars' construal of Hume's account is that the fatal ambiguity that crops up there is a direct byproduct of Sellars' taking vivacity to be

part of what we might call the *pictorial content* of an idea. Hume is an imagist. He argues that a mental entity represents what it represents in virtue of its being caused by what it represents and its *resembling* it.⁹⁵ So, for instance, my driver's license photo is a photo *of me* in virtue of the fact that there is a causal chain leading from me to the photo, *and* that photo resembles me. It is this latter condition that concerns us here. What it requires is that the photo have certain features in common with me. The color of the hair in the photo should be the same as the color of my hair, the relation of the eyes to the nose in the photo should be the same as the relation of my nose to my eyes, etc.

These features are what I call the *pictorial content* of the photo; Hume sometimes calls these the 'circumstances' of an idea.

When I shut my eyes and think of my chamber, the ideas I form are exact representations of the impressions I felt; nor is there any circumstance of the one, which is not to be found in the other. (T 1.1.1.3; SBN 2)

What is particularly noteworthy about the way that Hume treats these features of ideas is that he does *not* include among them the degree of force and vivacity of ideas and impressions. Earlier in this paragraph he notices,

the great resemblance betwixt our impressions and ideas in every other particular, except their degree of force and vivacity. (T 1.1.1.3; SBN 2)

Ideas resemble impressions in all their circumstances, but do not resemble impressions in their degree of force and vivacity. Therefore, force and vivacity must not be among the circumstances of an impression or idea. If, however, they are not among the circumstances of an idea, then they are not part of the *pictorial content* of the idea. They

⁹⁵This is Hume's story about singular representation. Of course, he also has a story to tell about general, or *abstract*, representation that is not as straightforward. Still, though, even in his story of abstract ideas, language, and custom there is an essential part played by singular representations that *are* images, and the *resemblance* between these.

are not features in virtue of which an idea represents what it does. In fact, this *has to be* a thesis that Hume endorses. Notice, for instance, the following series of claims.

Our ideas are copy'd from our impressions, and represent them in all their parts. When you wou'd any way vary the idea of a particular object, you can only encrease or diminish its force and vivacity. If you make any other change on it, it represents a difference object or impression. The case is the same as in colours. A particular shade of any colour may acquire a new degree of liveliness or brightness without any other variation. But when you produce any other variation, 'tis no longer the same shade or colour. (T 1.3.7.5; SBN 96)

Pace Sellars, Hume quite clearly believes that altering the force and vivacity of an idea *does not alter its content*, does not alter what that idea is an idea *of*. As Hume also clearly thinks that altering any *circumstance* of an idea does alter what that idea is an idea of, vivacity cannot be a circumstance of an idea.⁹⁶

What, then, is vivacity if not part of the pictorial content of an idea? Well, if we can be allowed for a moment to shift away from vivacity to force, an answer becomes apparent. The reason for this shift is because it is much less tempting to think of the *force* of an idea as part of its pictorial content. Rather, force would seem to be a matter of the *way* in which an idea with a particular pictorial content *acts* in the mind. (Hume calls this “the *manner* of our conceiving them”(T 1.3.7.4; SBN 96).) An idea is forceful just in case it cannot be shook, it *forces* itself on us, we cannot dislodge the idea from the theatre of our mind. A song stuck in your head, a haunting image, or a present visual sensation would all be paradigm examples of forceful ideas. Similarly, then, an idea is vivacious just in case it *jumps out at you*, it *catches your attention*, etc. That is, we can give both

⁹⁶It is worth noting reading the Copy Principle as Hume's principle of *intentionality*, reading it as determining what a perception is a perception *of* is a somewhat controversial position. This is at least because, when combined with the further (also somewhat controversially) Humean thesis that impressions are not copies of anything, this commits us to the position that only *ideas* have intentionality, impressions do not have intentionality, and ideas are all of impressions. I would argue that all of these are claims that Hume would make, but there is not space enough here to do so.

force and vivacity *functional*, rather than imagistic, readings. Force and vivacity are matters of the *behaviors* of ideas, not their content.⁹⁷

By making this move on Hume’s behalf, which he seems pretty clearly committed to anyway, we can now redraw our previous Figure 4 as follows.

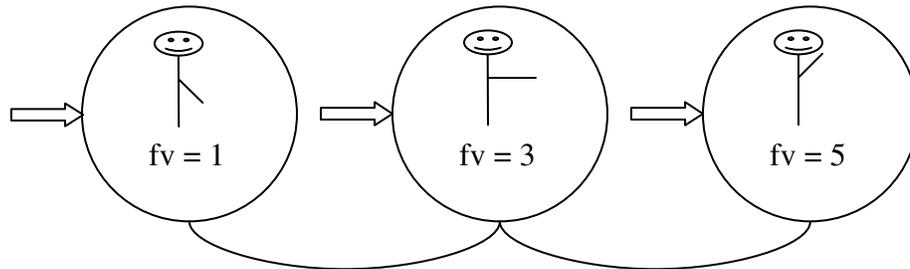


Figure 6

Here it should be clear that the *represented* object of each *representing* is not in any sense “faded”. The *representing* is either feeble and weak or forceful and vivacious, but this fact is now appropriately represented in our picture as an assignment of a degree of force and vivacity to each *representing*. The *representing* of the hand in the lowered position is the weakest, the *representing* of the hand parallel to the ground is slightly less weak, and the *representing* of the hand in the raised position is the most forceful and vivacious of all. The ambiguity that infected Sellars’ construal of Hume has been eliminated.

Furthermore, taking into account Hume’s contention that a memory becomes less and less forceful and vivacious the more time passes between its present occurrence in the mind and its original formation, we can now see one line of reasoning that Hume could pursue vis-à-vis the structure of complex ideas that represent temporally complex

⁹⁷This is not to say that ‘force’ and ‘vivacity’ are used *unambiguously* to mean only the manner in which ideas appear in the mind, even in Hume. It might very well be the case that Hume sometimes uses ‘force’ and ‘vivacity’ to speak of the manner in which ideas appear in the mind, and sometime uses them to speak of certain pictorial aspects of certain ideas. (E.g., that is a *vivacious* shade of red.) Rather the point here is just to make room for the former, manner-of-appearing, use of these terms. It is the possibility of using ‘force’ and ‘vivacity’ *this way, in addition to* the pictorial use, that give Hume’s theory a stay of execution here. So long as Hume is committed to *just* the pictorial use, he can safely use ‘force’ and ‘vivacity’ to describe features of *representings* without also using it to describe features of *represented*s.

states of affairs. The first step is to replace our Figure 6 with a new picture in which the coordination of degree of force and vivacity and the time at which an original impression is had has already been performed. That is, Hume can leverage coordination of force and vivacity with time of first appearance into a mental analogue of carbon-dating.⁹⁸ Thus we can picture the representation of a temporal sequence in Hume as follows.

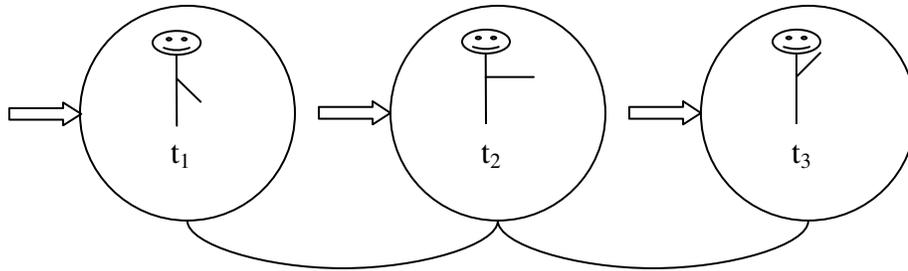


Figure 7

Here each representation comes with a time-stamp, reflected in its degree of force and vivacity. Of course, this correspondence of degree of force and vivacity and the time of formation of an idea is a vast oversimplification. Time of formation is only one of many things that affect the degree of force and vivacity of an idea according to Hume, and for the above procedure to really work, we would have to ignore all the others. Suppose we could do this, however. What would Hume's account of the representation of a temporal sequence be then, and how would it fare?

Well, we still have not had much to say about the connecting lines that we have imported into this Figure on Hume's behalf, about, that is, the structure into which representations are put in order to make them into representations of a temporal sequence. What I want to suggest now, however, is that this structure has become less important.

⁹⁸Of course, *real* carbon-dating only works because we have ways of *independently* corroborating its accuracy. Whether and what Hume's way of providing such evidence for the correlation of the degree of force and vivacity of a perception with the time at which the perception first appears in the mind is unclear at best.

Remember that the concern with structure earlier arose from a need to distinguish having a series of representations from having a representation of a series. Now that each representing in the series of representations is a representing of an event *at a time*, it seems more plausible to suppose that a series of *such* representings might very well, for all we have said, constitute a representation of a series. Such a series would represent an arm down at t_1 , an arm parallel to the ground at t_2 , and an arm raised at t_3 . Arguably, one cannot represent any of these *particular* times without *also*, at least implicitly, relating them to one another,⁹⁹ and so if Hume were right that we do the former on his account, we must also do the latter. Thus, the nature of the connections between these ideas as they occur *in our minds* (as, say, temporally successive, or even merely associated) would not matter as much as their *conceptual* relations to one another vis-à-vis their time stamps.¹⁰⁰

While this account of our capacity to represent temporal sequences may ultimately fail, and while it does not account for our ability to represent other kinds of complex states of affairs, the question with we are concerned is does it fail for reasons that Sellars has on offer, either in the lectures we have been examining, or elsewhere. What I want to suggest is that there is no objection to this account in the lectures. As we have seen, Sellars thinks that Hume cannot even get this far, that he has nothing to say on the matter at all. The inadequacy of Hume's account, however, cannot be dismissed so

⁹⁹Of course it is *Kant*, who is the most vocal advocate for the view that there is only *one* time, and that all *times* are a part of it. "For simultaneity or succession would not themselves come into perception if the representation of time did not ground them *a priori*. Only under its presupposition can one represent that several things exist at one and the same time (simultaneously) or in different times (successively)"(A31/B46). Of course, Kant puts this fact to a distinctly different use.

¹⁰⁰As Hume accounts for conceptual relations between ideas *via* the connections of ideas as they occur in our minds, things get complicated here. The thought, though, is that these ideas bear the special association relation to one another that is the *conceptual* association, as opposed to other ways of associating ideas.

easily. Rather it is due, firstly, to the fact that Hume's notions of force and vivacity are simply overburdened in his system, and so cannot be used to time-stamp impressions and ideas accurately. For instance, ideas can be *reinvigorated* by present impressions, giving them higher degrees of force and vivacity than they would have based solely on the time at which their original impression was formed. Furthermore, since a belief, on Hume's account, is a particularly forceful and vivacious idea, presumably evidence counting *against* a belief would weaken and enfeeble it, giving it *less* force and vivacity than would be appropriate to its time stamp. Etc. Hume has options available for dealing with such cases, but they are at least *prime facie* quite troubling.¹⁰¹

Secondly, Sellars offers yet another reason for rejecting Hume's account, not in the lectures we have been considering, but in *Science and Metaphysics*. There Sellars argues, on behalf of Kant that "what the representations of sheer receptivity are *of* is in no sense complex"(Sellars, 1967: 8), and that this is because, again following Kant, no *non-conceptual* representation can be *of* a complex. Now, it is unclear at this stage in our investigation just what being "conceptual" as opposed to "non-conceptual" entails, and which of these categories the representations for which we have seen Hume account properly fall under. The former topic, along with a defense of the above claims, will be the topic of the next section. We will return to the latter after that.¹⁰²

¹⁰¹For instance, Hume *could* hold that such cases merely make our representations of temporal sequences *false*, rather than voiding their status as representations altogether. He could also pursue a line on which these factors are, somehow, accounted, or compensated, for in the coordinating of degrees of force and vivacity with the time of the original impression formation.

¹⁰²It is also worth noting here is that while Hume's picture may seem a not entirely implausible way of representing a sequence of *experienced* events, it would not seem to be at all capable of speaking to how we go about representing a sequence of *fictional* events as such. That is, we are capable of representing temporally complex states of affairs as such even when they are merely imagined states of affairs, and in such a case it would seem entirely possible that each of the images that compose the parts of the fictional sequence would be equally forceful and vivacious. (They could all be as weak and feeble as any fantasy is,

We have now completed the first of the three tasks that we listed at the outset. We have seen that a certain criticism that Sellars makes of Hume—that Hume is unable to distinguish a forceful and vivacious perception from an impression of something forceful and vivacious—is unfounded. We have also, however, been able to glimpse a different criticism available to Sellars concerning the *structure* into which the elements of a complex representation are put to form representations of complex states of affairs as such. The second task on our list was to pursue this criticism, and as we have seen this means delving into the reasons behind the Sellarsian/Kantian claim that it is only *conceptual* representation that is capable of representing complex states of affairs as such.

Perhaps the best way to see why Sellars and Kant hold these views will be to see how he thinks we *do* represent complexes, and then to notice what the conceptual element in this account is, and what happens if we remove or replace it. To begin this task, we need to make a few distinctions. Firstly, Sellars’ contention on behalf of Kant that “what the representations of sheer receptivity are *of* is in no sense complex” can be taken in at least two ways.

1. The representations of sheer receptivity represent something, and this thing is simple.
2. The representations of sheer receptivity represent something, but do not represent it *as* complex (although what they represent may, in fact, *be* complex).

The way that Sellars puts his contention would make it seem that he intends the first of these: that he is saying something about *that which* the representations of sheer

on Hume’s account.) If this is right, however, it would seem that we are capable of representing temporally complex states of affairs without recourse to differences in the degree of force and vivacity of the component perceptions at all. Furthermore, it should be clear that the system outlined here applies *only* to *temporally* complex states of affairs, and there is not obvious way to extend this to complex states of affairs more generally.

receptivity represent. This, however, seems clearly to be a mistake. It is part of Sellars' project in the first chapter of *Science and Metaphysics*, from which this quotation is drawn, to show that what is represented by receptivity is *the same thing* as what is represented by the spontaneous understanding.¹⁰³ Both, in their own particular way, represent the spatio-temporal world, which is plenty complex. Rather, the point that Sellars makes here, and which we will soon see defended, is that the representations of sheer receptivity do not represent the spatio-temporal world *as* complex. Thus we find him a bit further on writing that “nothing represented by outer sense as such is a spatial complex”(Sellars, 1967: 8). It might be that what is represented by outer sense *is* a spatial complex, but it is not represented *as such* by *sense*, by receptivity.

So, Sellars thinks that what is represented by sheer receptivity is not represented *as* complex; receptivity cannot represent complex states of affairs as such. He does, however, think that what is represented by the spontaneous understanding is represented as complex, that it represents complex states of affairs as such.¹⁰⁴ Our task now is to investigate how he thinks it does this. The first step on the way to doing this is to notice with Kant that, “the cognition of every, at least human, understanding is cognition through concepts”(A68/B93). That is, when Sellars contrasts the representations of sheer receptivity with those of the spontaneous understanding, he is contrasting, following Kant, *conceptual* from *non-conceptual* representations. It is the former with which we will presently concern ourselves.

¹⁰³That is, both receptivity and the understanding are *first and foremost* faculties for representing *the world*. This is not to say that this is *all* that each can represent, or that each does not sometimes *misrepresent* the world. Rather, it is to claim that as Kant and Sellars see it, the primary function of each of the faculties—the function that makes any other uses of these faculties possible—is veridically representing the world.

¹⁰⁴He also thinks that there is a sense in which sheer receptivity can represent. More on this later.

Lucky for us then that Sellars tells us just where to find his answer to the question of how conceptual representations represent complex states of affairs as such, and of how these differ from non-conceptual representations of the same states of affairs (not as such).

If it is also true, as Wittgenstein held (*Tractatus*, 3.1432) that conceptual representations of relational states of affairs are to be construed as complexes of conceptual representations of their terms, the question obviously arises ‘What is the connection between the counterpart relations which bind conceptual representations of terms into conceptual representations of complex states of affairs, thus the conceptual representation that s_1 adjoins s_2 , and the counterpart relation which binds non-conceptual representations into non-conceptual representations of relational wholes, thus the impression of a green square adjoining a red square.’ I shall have something to say on this topic in Chapter IV. (Sellars, 1967: 26)

Notice that what the conceptual representations represent are complex states of affairs, while what the non-conceptual representations represent are relational wholes. Notice also that already we see that the *way* that conceptual representations operate in representing complex states of affairs is by standing in a certain kind of *relation* to one another. These will both be important points, for our further discussion. For now, though, we must follow our lead to Chapter IV.

The relevant place to pick up this thread is the following long paragraph in which Sellars describes how he thinks we conceptually represent complex states of affairs as such.

Now what Wittgenstein established was that whether one does it perspicuously or not, one can say of two objects that they stand in a certain relation by placing the corresponding referring expressions in a counterpart relation. Thus we say

a is larger than b

or

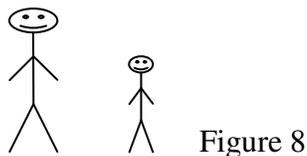
a_b

in either case what we have done is form an expression which, from the standpoint of its semantical functioning, is a dyadic configuration of the names '*a*' and '*b*'; it is, in other words, an

$R^* [\cdot a, \cdot b]$

where this is a common noun which applies to items consisting of an $\cdot a$ and a $\cdot b$ related in a way which does the job, in the language to which it belongs, which is done in our subject-predicate language by concatenating these names with a predicate. (Sellars, 1967: 109)

Here Sellars picks up just where he left off in Chapter 1 with the point he takes Wittgenstein to have proved, namely, that we represent complex states of affairs as such by placing counterparts of the constituents of those states of affairs into relations which are the counterparts of the relations to which those constituents stand. So, for instance, consider the following complex state of affairs, in which the figure on the left is larger than the figure of the right.



Sellars' suggestion is that the way we represent this state of affairs is by placing the names *a* and *b*, which respectively refer to the figure on the right and the figure on the left, into a certain relation *with each other*, a relation that is the counterpart in our language of the relation in which the figures stand to one another. In our language, the counterpart relation in which we would place the names *a* and *b* would be placing them on either side of the words 'is larger than' (or a suitable functional analogue of 'is larger than').

This parenthetical remark brings us to a crucial point. It will, no doubt, have been noticed that in the above quotation the names *a* and *b* appear dot-quoted, and this fact requires an explanation. Dot-quotes, as Sellars uses them, provide a way to mention bits

of language *functionally classified*. So, while ‘red’ mentions the English word ‘red’, ‘red.’ mentions any bit of language, in any language, that plays the role that ‘red’ does in English. Thus, its scope includes the German ‘rot’, the Spanish ‘rojo’, the French ‘rouge’, etc. When Sellars, then, asserts that placing an *·a·* and a *·b·* in a certain relation R* represents *a* and *b* as standing in a certain relation R, he means that placing *any terms that are functionally equivalent to ‘a’ and ‘b’* in relation R*—or *its* functional equivalent—represents *a* and *b* as standing in relation R to one another. We can say, for instance, that John is larger than Judith, or that Johnny is bigger than Judy, etc. so long as each member of the pairs ‘John’ and ‘Johnny’, ‘Judith’ and ‘Judy’, and ‘is larger than’ and ‘is bigger than’ play the same role in the language as one another.

It is important to note here that while Sellars argues that it is in principle possible—and perhaps most perspicuous—to eliminate predicative *expressions* from a language (because, as he also argues, predicative expressions do not *refer*), we cannot eliminate the *role* that such predicative expressions play (properly construed, of course).¹⁰⁵

Let me emphasize that this is not to say that singular statements could in principle, ‘consist of referring expressions’ in the sense of being a mere list—the crudest possible form of nominalism. It is rather the more subtle point that one says *how* objects are by inscribing or uttering the corresponding referring expressions in a certain manner. (Sellars, 1967: 109)

¹⁰⁵We must tread very carefully here, for Sellars does claim that “not only are predicative *expressions* dispensable, but the very *function* performed by predicates is dispensable”(Sellars 1980: 59). As is clear from the quotation in the main text, he does not mean by this that there could be a language that consisted merely of names. Rather he means that predicates do not themselves represent anything, and so have no *semantic* function, strictly speaking. They still play a role, however, in *structuring* language. In particular, they are ways of structuring names, which structured names represent the ways that worldly objects are.

The job of predicates—be these predicative *terms* or predicative *relations*—is to allow us to say “*how* objects are”. That is, it is to allow us to represent *complexes*.¹⁰⁶ Thus, while we could eliminate the sign-design ‘is larger than’ from our language and continue to represent one thing being larger than another by, for instance, placing the name of the larger thing above the name of the smaller, here that *placement* would be doing certain work in our language. Placing names in *that* relation, rather than the relation in which they are on either side of an ‘is larger than’, says of the things that one is larger than the other. The importance of this is that it allows us to see that there is a functional role played in our language by ‘is larger than’s, that is, we can speak of ‘is larger than’s. What we must turn to now is the nature of this functional role.

Sellars is quite explicit and clear on what the role is that any linguistic item must play in order to be have this kind of function. He believes that this is a broadly *inferential* role (although not every relevant connection between linguistic items is an inference). In fact, Sellars gives a list at the close of the chapter that we have been examining of what kinds of relations constitute inferential roles. These are his well-known language-entry, language-language, and language-exit moves.¹⁰⁷ They are essentially kinds of rules for connecting judgments, sensations, and volitions. Names and predicates are the kinds of linguistic items that license and prohibit such connections. For instance, the proper use of

¹⁰⁶In saying that it is predicates that allow us to represent complex states of affairs as such, I do not mean to imply that predicates *themselves* represent any kind of object. Rather, predicates allow us to position our representations of objects in an inferential structure that, in turn, represents these objects as connected to one another in determinate ways. (It is worth noting that this is the case even with monadic predicates. The role of such predicates, like other, is to place the name of an object into an inferential structure, i.e., into inferential relations with other objects.)

¹⁰⁷It is important to note here that language-entry and language-exit moves are *not* themselves inferences. On pain of robbing Sellars’ system of crucial parts of its epistemology, philosophy of perception, etc., inferences are strictly intra-linguistic moves. Still, language-entry and language-exit moves do contribute essentially to the inferential role of a linguistic item. That is, linguistic items are often, if not always, defined in part by their relation to certain language-entry and language-exit transitions.

a ·lightning now· authorizes the use of a ·thunder soon· in virtue of the material implications that connect these two judgments, which implications represent the *relation* of lightning to thunder.

Similarly, Kant is often read as a kind of inferentialist.¹⁰⁸ Considering the concept of a body, for instance, he writes,

Thus the concept of body *serves as a rule* for our cognition of outer appearances by means of the unity of the manifold that is thought through it. However, it can be a rule of intuitions only if it represents the necessary reproduction of the manifold of given intuitions, hence the synthetic unity in the consciousness of them. Thus in the case of the perception of something outside of us the concept of body *makes necessary* the representation of extension, and with it that of impenetrability, of shape, etc. (A106. emphasis mine)

Kant, of course, does not mean that whenever we employ ‘body’ in a judgment, we must also employ ‘impenetrability’, ‘shape’, etc. in further judgments. What he means is that the concept of a body serves as a *rule of inference* according to which if one is committed to something’s being a body, then one is thereby also committed to its being impenetrable, to its having a shape, etc. Analogously to Sellars, Kant thinks that placing intuitions into certain inferential relations with one another is how we represent the manifold of intuitions as being a certain *way* (of containing certain necessary relations).

The importance of inference for Sellars and Kant in the current context—determining why they believe only conceptual representations can be of complexes—cannot be underestimated. Consider again the two figures pictured above.

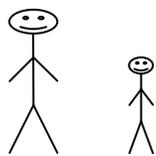


Figure 8 (repeated)

¹⁰⁸Most prominently, of course, by Sellarsians themselves. Cf. Rosenberg ,(2005); Brandom, (1994).

According to Sellars, the way to represent the figure on the left being larger than the figure on the right is to place the names of these two figures into a certain relation with one another; in English we place their names on either side of an ‘is larger than’. Now, it should be clear that these two figures stand in *many* relations to one another other than the left’s being larger than the right. They are *next to* one another, they *resemble* one another, etc. What explains, therefore, that placing the names of each of these figures on either side of an ‘is larger than’ represents the one as being *larger than* the other, rather than say one’s be next to the other, or one’s resembling the other? The answer, according to Kant and Sellars, is that it is the specific inferential role that ‘is larger than’s play in English that does so.

That is, one predicate (or name) will be differentiated from others by the inferences it licenses, prohibits, etc. in a language. For example, ‘a is larger than b’ and ‘b is larger than c’ licenses ‘a is larger than c’, but ‘a resembles b’ and ‘b resembles c’ does not license ‘a resembles c’. Furthermore, pace Wittgenstein, Sellars (and Kant) argue that not only are there *formal* rules of inference, but there are also *material* rules of inference.¹⁰⁹ So, for instance, ‘a is to the west of b’ licenses ‘b is to the east of a’, and ‘lightning now’ licenses ‘thunder soon’. It is these material inferences that allow for the individuation of linguistic-functional items. Were terms to have their meaning merely in virtue of the role they played in formal inference, most predicates would be indistinguishable; we would not be able to represent particular complexes *determinately*.

This, it turns out, is the core of the Kantian-Sellarsian critique of Hume. Sellars describes Hume’s position vis-à-vis material inference as follows.

¹⁰⁹Cf. Sellars (1953).

Those who take this line claim that “It is raining, therefore the streets will be wet”, when it isn’t an enthymematic abridgment of a formally valid argument, is merely the manifestation of a *tendency to expect* to see wet streets when one finds it raining, a tendency which has been hammered into the speaker by past experience. In this latter case it is the manifestation of a process which at best can only *simulate* inference, since it is an habitual transition of the imagination, and as such is not governed by a principle or rule by reference to which it can be characterized as valid or invalid. That Hume dignified the activation of an association with the phrase “causal *inference*” is but a minor flaw, they continue, in an otherwise brilliant analysis. (Sellars, 1953: 315)

Sellars here describes the Humean position as one that substitutes *mental associations* for material inferences. Since among the work we have been supposing that material inference does for Sellars is to individuate names and predicates, the question then becomes: can mental association be used to individuate names and predicates?

The answer to this question must be a resounding ‘No’. The associations one makes between certain ideas—or linguistic items—in one’s mind are notoriously fickle, while the content that is represented by those ideas is, for the most part, fixed. Associating a dog-idea with a cat-idea does is not the same as representing dogs and cats as being of the same kind. One may simply be inclined to think of a dog whenever one thinks of a cat, while remaining perfectly clear that dogs and cats are entirely different kinds of animals. Furthermore, if one did take dogs and cats to be the same kind of thing, one would be *wrong* to do so. There is a dimension of normativity in our concepts for which association simply cannot account.¹¹⁰

If Sellars is right that we represent complex states of affairs by placing names into relations with one another that are the counterparts of the relations in which the referents of these names stand (a big if), then it is certainly more plausible to think that these counterpart relations will be *inferential* rather than *associative*. There are only so many

¹¹⁰Notice that this much is admitted by even the most robust naturalist about concepts (e.g., Fodor, (1990)).

associations one can make, or even be disposed to make, and these are clearly inadequate for representing all the many relations that things can stand in to one another. There could be, on the other hand, exactly as many inferential relations, especially when one includes material-inferential relations, as there could be relations among objects. In fact, the whole *purpose* of placing names into inferential relations, on Sellars' view, is to create an adequate picture or map of the world (the sum total of complex states of affairs), and so on this view, we strive to accomplish just this kind of one-to-one correspondence of inferential relations to complex states of affairs.

For this picture-theory to work, it is essential that the complex representations involved be *unambiguous* pictures of their objects. *Associating* a mental representation of dog with a mental representation of a cat is not sufficient for producing an unambiguous representation of, for instance, a dog *next to* a cat. The normativity of inferential relations helps here because with it we can place these representations (of the dog and of the cat) in more determinate relations with one another. Instead of merely associating these two mental items, we can place them into relation with, say, a *third* mental item, an 'is next to', which signals their places in an unambiguous system of inference. It is the normative aspect of inference which takes us beyond what we *associate* a mental representation with, to what we are, say, *committed to*, or *forbidden from*, or *permitted to*, given that we have certain mental representations. These further resources are what provide for the disambiguation of mental representations.

This, then, is ultimately the source of the Kantian-Sellarsian critique of Hume's account of the representation of complexes (including, but not limited to, spatial and temporal complexes). To represent a complex state of affairs requires a set of

representings that can themselves be *determinately* classified as having that particular state of affairs as their content. Hume's account is inadequate because the only means by which he can classify such representings is by their associations, and associations are themselves inadequate to this task. For Sellars, a complex of representings, such as a_b , can come to represent a complex state of affairs only if the former complex stands in inferential relations with other complexes, which relations give it a determinate content. For Hume, on the other hand, the only relations that complexes of representings can stand in are relations of association, and relations of association are not sufficient for giving these complexes of representing any determinate content.

It may be here objected that this critique of Hume falls short insofar as it takes for granted that the only way that Hume has to fix the content of a complex of representings is via an analogue of placing those representings in material inferential relations. This, however, is a thesis to which Hume would object. It is not, Hume would contend, by associating representings with one another that the subsequent representing has the content that it does. Rather, it has this content because the complex idea formed by association itself *resembles* its content. We have, the objection goes, been giving short shrift to Hume's actual theory of content.

The response to this objection, however, has already been given earlier in our discussion of the representation of a temporal sequence. Taken strictly, this suggestion would entail that we represent a temporal sequence by placing representations into temporal sequences. What we saw earlier, however, is that nothing in such a sequence of representings would itself represent a temporal sequence.

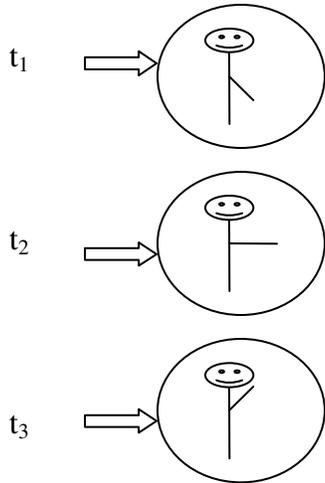


Figure 2 (repeated)

That is, none of the ideas had at t_1 , t_2 , or t_3 would themselves represent anything complex, and if it is the whole of the sequence of representings that is to represent the whole of a sequence of events, again the burden falls on association to be the tie that binds this sequence of representings together.

We can now see the sense in which Sellars' criticism of Hume in his lectures is appropriate. While it is not the case, on the reading of Hume that we have presented, that Hume confuses a vivacious idea for an idea of something vivacious, it is still the case that Hume runs into a vicious ambiguity. It is that between the representation of one complex, and the representation of an entirely different one. The ties that bind together representations into a picture like this,

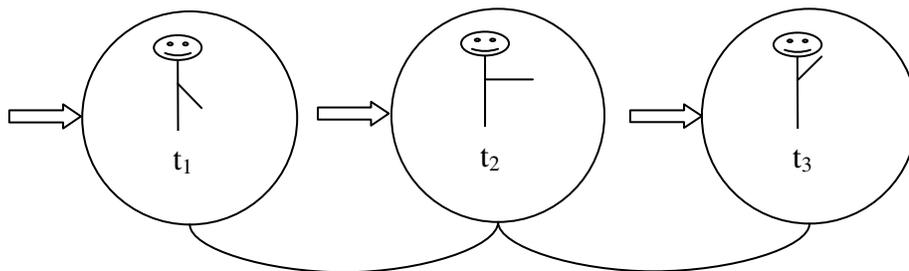


Figure 7 (repeated)

are not enough to distinguish, for instance, a representation of person raising his hand from a representation of three people each with their hands in different positions. These ties are mental associations, and, appropriately, as in Figure 2, there is no representation of the entirety of the temporal sequence.

Despite the importance that Sellars and Kant both place on the conceptual aspects of representation, both also maintain that there is a sense in which sheer receptivity, or the non-conceptual part of experience, provides its own representations. As we have just seen, neither can hold that receptivity can represent a complex state of affairs as such, but there is for both a sense in which it represents nonetheless. This is brought out most clearly when we focus on the distinction that Sellars points to in Kant between *sensation* and *intuition*. Here is Kant's explication of the genus representation and the places of sensation and intuition in it.

The genus is representation in general (*repraesentatio*). Under it stands the representation with consciousness (*perceptio*). A perception that refers to the subject as a modification of its state is a sensation (*sensation*); an objective perception is a cognition (*cognition*). The latter is either an intuition or a concept (*intuitus vel conceptus*). The former is immediately related to the object and is singular; the latter is mediate, by means of a mark, which can be common to several things. (A320/B377)

A sensation is a representation, with consciousness, that refers to the subject as a modification of its state. That is, a sensation is an apperceptible mental state that does not purport to be *about* anything.¹¹¹ We will come, in a moment, to the question of how this is compatible with it nonetheless representing something. An intuition, on the other hand, is a—get ready—singular, immediate, objective representation with consciousness. That

¹¹¹I take it that when Kant says that a sensation *refers to* the subject as a modification of its state, he means 'refers' here in the way that we do when we speak of, for instance, referring someone to a doctor. That is, sensations point towards, or are an indication of something in, the state of the perceiver, but they are not *about* the state of the perceiver. It is certainly *not* Kant's position that sensations are all *of* the oneself. Such a position is one that Kant takes, for instance Descartes to hold, and is one that he seeks to correct.

is, it is an apperceptible mental state that does purport to be about something, that is about just one thing (as opposed to a concept, which can apply to many things), and that is immediately related to its object (again, as opposed to a concept, which relates only mediately to its object, via being combined with an intuition in a judgment).¹¹²

The difference between a sensation and an intuition with which we will be concerned is that the latter but not the former purport to be about something. As we have seen, intuitions do this by standing in certain inferential relations with one another; intuitions are in this sense *conceptual*. It is for this very reason that sensations are *non-conceptual*. Sensations do not stand in inferential relations to one another. They never enter into judgments. They are *mere* mental states.

We must wonder, then, in what sense these mere mental states *represent*, and what role they play in cognition. It by addressing the latter that we can come to address the former. As Sellars sees it, sensations are not something with which we simply *find* ourselves, but rather they are something that we *postulate* as an explanation of certain mental phenomena, specifically, as a way of explaining our *conceptual* responses to non-conceptual stimuli, i.e., the world. He calls this postulation the sense impression inference, and writes of it that

[I]ts primary purpose is to explain the occurrence of certain *conceptual* representations in perceptual activity. [...] Thus, the sense impression inference is an attempt to account for the fact that normal perceivers have *conceptual* representations of a red and rectangular object both

- (a) when they are being affected in normal circumstances by a red and rectangular object; and
- (b) when they are being affected in abnormal circumstances by objects which have other, but systematically related characteristics. (Sellars, 1967:17)

¹¹²Returning to our earlier discussion of Sellars, intuitions are *like* names in that they stand for objects; concepts are like predicates in that they provide for *ways* of relating intuitions to one another so that the complex formed represents the world.

On Sellars' view, we postulate sensations as a way of explaining why we respond—correctly and incorrectly—with conceptual representations to various non-conceptual stimuli.¹¹³ The sense-impression inference is an *explanatory* postulate meant to account for the general success and occasional failure of perceptual cognition.

Of course, as an explanatory hypothesis, the sense impression inference cannot merely be that there is *some* feature of experience that explains perceptual cognition. Rather, if it is to be explanatory, it must postulate that there is a particular feature of experience with certain qualities that explain how just this feature can do just this work. As Sellars puts it vis-à-vis the sense impression inference that explains color-concepts,

Thus, these non-conceptual states must have characteristics which, *without being colours*, are sufficiently analogous to colour to enable these states to play this guiding role. (Sellars, 1967, 18)

Sensations cannot be colored. (Anything colored is extended; no mental item is extended; sensations are mental items.¹¹⁴) However, if sensations are to explain why we respond with color-concepts to colored physical objects, they must be in some sense analogous to colored physical objects. They must share a *structure* with colored physical objects. That is, they must behave in certain ways that explain both their correlation with physical objects of various colors and our conceptual responses to these. If the sense impression inference is to be explanatory, it must postulate a manifold of sensations that shares a structure with physical objects—in this case, the structure of colors.

¹¹³This is the *primary* reason for making the sense impression inference. Once made in this case, however, it also becomes appropriate to make in the case of, say, animals who do not respond *conceptually* to anything, but who clearly *represent* complexes nonetheless.

¹¹⁴This is, for Sellars, an argument that can only be given in the Manifest Image in which colors are properties of physical objects. In the Scientific Image, for Sellars, it will turn out that, in fact, it is *only* *sensa* that are colored.

Returning to our original concern—with spatio-temporal complexes—we can extend the sense impression inference to this realm as well. We can, and Sellars does, argue on Kant’s behalf that just as Space and Time are the form of outer *intuition* (objects of possible experience all exist within a single spatio-temporal framework), there must be some *analogously* structured manifold of sensation that explains our (appropriate and inappropriate) conceptual responses to encounters with such objects. (Kant writes about a duration that is not time, for instance.) Sellars designates the analogues of space and time in the manifold of sensation as σ and τ respectively, and points out that as in the case with colors, σ - and τ -structured manifolds are not themselves spatial or temporal, but must be structured in a way closely analogous to the way that space and time are.

It is this manifold of structured *sensations* with which Sellars takes Hume to be concerned, and which he argues cannot represent anything complex as such.¹¹⁵ The best we could muster on Hume’s behalf by way of an account of how we represent spatio-temporal complexes was a manifold of non-conceptual representations structured in a way that mimicked the structure of their collective object. This is exactly what Sellars takes the manifold of sensation to be, and as we have seen, neither he nor Kant think that this is sufficient to the task at hand. Attempts to account for representations of complexes in terms of such a manifold do not have the resources to make such representations determinate in the way they would have to be to succeed. Without such determinateness, the manifold of sensations is simply inadequate to the task of representing complexes as such.

¹¹⁵Of course, Hume is concerned with this aspect of mental representation because he thinks it can do the work involved in representing complexes as such. Clearly, Sellars’ thinks that this Humean thesis *about* sensations is a mistake.

These indeterminacies, however, do not affect the role that Sellars provides for the manifold of sensations. What Sellars requires of the manifold of sensation is just that it be structurally analogous to the spatio-temporal world. He is willing from the outset to grant Hume this much. The problem with Hume, Sellars argues, is that he takes this structural analogy to be sufficient for determinate representation. Sellars sees that it is not, and replaces it with a further level of representation—conceptual representation—that he argues is sufficient. Still, Sellars, along with Kant, does refer to the manifold of sensation as a kind of *representation*, and we are now in a position to see why this is so.

Remember that Kant classifies sensations as representations that refers to the subject as a modification of its state. A sensation, that is, does not purport to be *about* anything, and in particular does not purport to be about objects in space and time. How, then, we must wonder, does it *represent* objects in space and time? The answer seems to be that the sense in which sensations stand for objects is that they stand *in* for objects.¹¹⁶ They represent objects, not in the sense that a portrait represents a person, but rather in the sense that a lawyer represents his client—by doing work on behalf of the client that the client cannot himself do. Here, the work that is to be done is providing the proximate causal antecedent of conceptual behavior. The fact that we make *mistakes*, demonstrates that objects cannot do this work; if they did such mistakes would be inexplicable. Rather, objects enter into the story only *mediately*, through their mental *proxy*, sensations. Sensations stand in for objects in our mental lives, and by doing so provide the explanation of how it is that our conceptual behavior comes to generally track, but sometimes mis-track, worldly objects. This is because sensations are imperfect stand ins

¹¹⁶Cf. Rosenberg, (1986).

for worldly objects, sometimes presenting the analogue of, say, a red object, when the world contains, say, a white object in red light.

Notice that it is not just non-conceptual representations that stand in for worldly objects; i.e., Kant and Sellars do not use ‘representation’ differently when considering sensations and intuitions, respectively. For both Kant and Sellars, intuitions too stand in for worldly objects. As we saw earlier, they both share the view that it is by placing intuitions into inferential relations with one another that we come to represent complex states of affairs. Here the intuitions represent in virtue of their beings stand-ins for the objects they represent. Of course, there is a crucial difference between the way that intuitions represent and the way that sensations do, which allows for intuitions but not sensations to represent complexes as such. As we have seen, the former kind of representational standing-in is *conceptually* structured, whereas the latter is non-conceptually structured, in particular is σ - and τ - structured instead.

The manifold of sensations, then, represents the world of objects in space and time by standing in for this world in our mental lives. We can, with Hume, agree that this is accomplished by that manifold being structured in a way that mimics the structure of the spatio-temporal world (although it is not clear that this is achieved by *association*), while we can still maintain, pace Hume, that *this* kind of representation is inadequate to representing the spatio-temporal world *as* complex. It may be that each piece of the manifold of sensation represents some piece of the spatio-temporal world, but it is not the case that anything in that manifold represents anything in the world *as* related to anything else. This, we can follow Sellars and Kant in saying, is only possible via conceptual

representation. Thus is the Kantian critique of Hume's account of the human ability to represent complex states of affairs completed.

Chapter Five

Hegel's Account of Rule-Following

At the core of Kant's *Critique of Pure Reason* is the Transcendental Deduction, the argument in which Kant attempts to justify the use of pure *a priori* concepts of the understanding, or Categories. This argument is a practical one.¹¹⁷ It begins with what Kant takes to be a necessary goal of every creature like ourselves: to conceive of oneself as a single, unified subject of experience persisting through time.¹¹⁸ Kant argues that the only means to achieving this goal is to unite the manifold of intuitions in a single cognition, that is, to have a single thought the content of which speaks not just to how the manifold is at the time of the thought, but also how it must be at other times, in other locations, and in certain other counterfactual situations. This, in turn, he argues, is only possible via the employment of the Categories in correct object-concept employing judgments. Thus Kant's argument goes from the need to conceive of ourselves as single, unified subjects of experience to the justification of the use of the Categories.

Of course, along the way, Kant takes on board a good deal besides these two features of experience. Most significantly, for reasons concerning the role of judgment in

¹¹⁷In casting the Transcendental Deduction as a piece of practical reasoning I am following Sellars, (1964) and Rosenberg, (2005).

¹¹⁸According to Kant, *we* are the types of creatures that passively receive sensory manifolds over time, use concepts to unite these manifolds, and can apperceive our doing so. Thus the goal listed is a *constitutive* goal; it is one that we must all meet in order to *be* the kinds of creatures we are.

uniting the manifold of intuitions, and with the unity of judgment itself, Kant becomes an inferentialist about conceptual content.¹¹⁹ One central purpose of this paper is to explore the consequences of Kant's claim that these issues require us to be inferentialists, and to conceive of concepts as *rules* for uniting the manifold of intuitions. First, however, it will be worth our while to take note of another thesis advanced by Kant in the course of this argument, namely, that the experiencing subject must be able to undertake the making of correct object-concept employing judgments as a means to an end.

It is significant that this thesis concerns ends that the experiencing subject has and means that he takes to satisfy them. This is particularly important because for a piece of practical reasoning to function as such, the end and means that it describes must themselves be such that the subject of the argument can *correctly represent to himself* what it would be to have such goals and take such ends. A goal is only someone's goal if that person can know in what it would consist to meet that goal. A means is only someone's means if that person can know in what it would consist to take those means. The purpose of practical reasoning is to structure the goal-oriented behavior of the one who engages in it. When one is the subject of a piece of practical reasoning—insofar as one is rational, and the reasoning is sound—one *has* the goal and *adopts* the means prescribed by that argument. I.e., the result of a piece of practical reasoning is the forming of an *intention*. One can only form an intention, however, if one can *understand* the goals and means that are to be the object of that intention.

The significance of this for the current investigation is that whatever it is to correctly employ object-concepts in judgments—that is, on Kant's view, whatever it is to

¹¹⁹Kant is an inferentialist of a very particular type. He is crucially different from, say, Brandom insofar as he makes it a necessary condition on something's being a concept that it bear inferential relations not only to other concepts, but also crucially to *intuitions*. I.e., he places language-*entry* moves at the fore.

follow a rule—must be the sort of thing one can form the intention to do. It follows from this that following a rule must be something *recognizable as such* to someone who is doing so.¹²⁰ We can call this the Internalist Premise. As we are about to see, this is a crucial background assumption in Hegel’s arguments concerning rule-following.¹²¹

The first three chapters of Hegel’s *Phenomenology*—gathered under the heading “Consciousness”—are concerned with non-inferentialist theories of representation, and their frequent ally, correspondence theories of truth. Hegel’s procedure in these chapters is to demonstrate of each theory he considers that it is what Kant would call transcendent. That is, he demonstrates that each theory violates a more general version of the Internalist Premise by accounting for correct representation (and knowledge, truth, etc.) in ways that make it in principle unrecognizable as such to the representing subject.¹²² Since we are beginning with Kant and *inferentialist* theories of conceptual representation, we do not have to delve into these sections of the *Phenomenology* in any detail. We can instead pick

¹²⁰It must be recognizable as such that one is following *some rule or other*, that this rule is one governing the employment of an *object*-concept, and that one is following it correctly. It may still be unavailable to one just which *particular* rule one is following in doing these, which particular object-concept one is employing.

¹²¹It is worth noting that this premise is one which, although it is fairly clearly at work in many of the contemporary discussions of rule-following, is rarely given a defense. This is true despite the fact that it is perhaps the very premise at the heart of the disagreement between social-inferentialists (e.g., Brandom) and so-called Meaning Externalists (e.g., Dretske, Fodor, Millikan). Of course, the current defense in terms of practical reasoning may be as controversial as the Internalist Premise itself—a meaning externalist might very well want to be an externalist about ends and means as well—but it at least attempts to shed some light on *reasons* one might have for holding the Internalist Premise. Gareth Evans is another philosopher in the business of giving such reasons. Cf. Evans, (1982).

¹²²Forster, (1998) argues that the opening chapters of the *phenomenology* (along with certain chapters in the section on Reason) do work towards *two* important Hegelian theses: that conceptual understanding is necessarily linguistic, and that language is necessarily communal. On the present reading, it is only the former thesis that is engaged in the sections on Consciousness, despite the fact that, once the latter thesis has also been adopted, Hegel can further say of Consciousness that it fails for neglecting *it* as well. We are about to see how Hegel’s argument that language, and therefore conceptual understanding, develop in the course of his treatment of Self-Consciousness.

up Hegel's dialectic at the point that he himself takes the Kantian turn—an approach to the *Phenomenology* that Hegel himself would find quite appropriate.

Hegel begins his investigation into rule-following with what *seems* like a declaration, not of inferentialism, but of *idealism*.

In the previous modes of certainty what is true for consciousness is something other than itself. But the Notion of this truth vanishes in the experience of it. What the object immediately was *in itself* [...] proves to be in truth, not this at all; instead, this *in-itself* turns out to be a mode in which the object is only for another. (PS §166)¹²³

As self-consciousness, it is movement; but since what it distinguishes from itself is *only itself as itself*, the difference, as an otherness, is *immediately superseded* for it; the difference *is not*, and *it* is only the motionless tautology of: 'I am I'; but since for it the difference does not have the form of *being*, it is *not* self-consciousness. (PS §167)

One *might* here take Hegel to be asserting a kind of idealism. He *seems* to be saying, especially in the second passage, that self-consciousness is aware not of anything outside of itself, such as an external world, but only of itself and its own activities—“what it distinguishes from itself is only itself as itself.” Furthermore, on this reading what self-consciousness takes to be outside of itself, the external world, “is immediately superseded for it,” and the difference between the two “is not”. Returning to the first passage, this reading would have it that there is no way that the external world is apart from our conceiving it; “this in-itself turns out to be a mode in which the object is only for another.” This all certainly *sounds* like idealism.

In fact, however, what Hegel is here noticing is not (yet) idealism, but rather only a somewhat more mundane fact about the inferentialist's take on meaning: that when we specify the meaning of a given term, we do not do so by pointing to some worldly object

¹²³All citations from the *Phenomenology* are taken from Hegel, (1977). I will continue to abbreviate this PS, followed by section number.

for which that term stands, but rather we assign that term a place in *our* language. This claim was perhaps made most clearly and explicitly by Sellars via his account of the meaning rubric and his accompanying theory of meaning.¹²⁴ Consider the claim that,

(1) 'dog' refers to dogs

On a certain kind of referential account of meaning, this sentence concerns the relation of tokens of certain linguistic items, DOGs, to certain objects in the world, dogs. It concerns a word-world relation. For the inferentialist, this will not be so. Because the inferentialist accounts for the content of conceptual representations in terms of the *inferences* in which a term figures, this content will consist, instead, of various word-word relations.¹²⁵ So, using the Sellarsian analysis of the meaning rubric, the inferentialist will construe (1) as

(2) *dog*s are ·dog·s.

Here *dog* picks out the concept >>dog<< in some target language (here English), and says of it that its inferential role is the same as that of the >>dog<< *in the language used by the person making this statement* (again, English).¹²⁶ The inferentialist claims that statements about the meaning of a concept are always given in terms that relate the inferential role of that concept to the inferential role of a concept in actual use by the speaker. This is because, since meaning is *not* understood by the inferentialist as a word-world relation, there is nothing *outside* of word-word relations—properly understood—that *can* be an articulation of the meaning of a concept. On an inferentialist account, that

¹²⁴Cf. Sellars, (1962).

¹²⁵In fact, for Kant it will consist in certain language-*entry* as well as language-language moves. It is a short step to adding language-*exit* moves to this list. Cf. Sellars, (1954). As language-entry and language-exit moves are both, *in some sense*, word-world relations, one must be very careful here about oversimplifying the inferentialist's position. Thus, the step from inferentialism to idealism is not as straightforward as it might at first seem.

¹²⁶I am here using Continental quotation marks, '>> <<', to name inferentially articulated concepts.

is, meaning cannot be reduced to reference (as reference is understood by, for instance, a meaning *externalist*).

Perhaps surprisingly, *this* is Hegel's point in the above quotations. In the previous "modes of consciousness" that Hegel has considered—accounts of conceptual representation that construed representation as consisting in a word-world relation—"what is true for consciousness is something other than itself". On such accounts, representations have the content that they do because they bear a certain relation to something *in the world*, and thus to something distinct from the one making such a judgment. The *truth* of a judgment made by employing such representations thus depends on the *judgment's* standing in a certain relation to the world. What the inferentialist claims, on the other hand, and what Hegel takes himself to have shown, is that "this *in-itself* turns out to be a mode in which the object is only for another." Hegel's contention is that the content of a conceptual representation consists in its relation to other concepts that the one using the concept could himself also employ. The *content* of a conceptual representation is not a thing in the world that bears a relation to the tokens of that representation, but rather is the inferential role that that concept plays in the language of the one who uses it.

Hegel's next move, then, is to explore the nature of this language-user. If all reference is *intra-linguistic*—in the sense that it is not a word-world, but word-word relation—what needs investigating is the nature of such intra-linguistic relations. Recognizing, with Kant, that the conceptual rules that govern such relations are underdetermined by sensory inputs, Hegel, like Kant, concludes that conceptual rules are crucially tied to the spontaneous understanding of the experiencing subject. That is,

because sensory inputs underdetermine the conceptual rules that one follows in organizing one's experience, one is, at least in this sense, free to use whatever concepts one wishes—so long as they accord with the Categories. Hegel's first supposition—which he later rejects—is that this is *all* that there is to such rules: i.e., that the experiencing subject is completely unhindered in what rules he chooses to follow and in what the content of these rules consists.

Hegel calls this conception of the rule-following subject Desire and its object Life. We can, for present purposes, set aside the question of why he chooses just these terms.¹²⁷ What is more important is understanding Hegel's picture of the interaction between Desire and Life. Of Life, he writes,

Thus the simple substance of Life is the splitting-up of itself into shapes and at the same time the dissolution of these existent differences; and the dissolution of the splitting-up is just as much a splitting-up and a forming of members. (PS§ 171)

The idea here is that according to this conception of rule-following, the nature of the objects that one represents using conceptual rules depends, in some sense, on the conceptual rules one uses to represent them. Again, there is supposed to be, on this account, nothing more to the fact that *x* is *F* than that an experiencing subject takes *x* to be *F*. This is because the meaning of 'F' will always be whatever the experiencing subject determines it to be. So, if that subject judges that *x* is *F*, then it follows that *x* is *F*. This is not because the experiencing subject changes the world so as to make *x* *F*, but rather because he is free to change the meaning of the term 'F' so as to include *x* in its

¹²⁷Pippin, (1989) suggests rightly, I think, that Hegel's choice of "Desire" signals both the pragmatist/internalist strain that Hegel initiates in this section of the *Phenomenology*, according to which what counts as correct and incorrect for a subject will be a matter of what his *interests* are in the subject matter, and that since the subject conceives of the external world as something *to be known*, it is conceived by him essentially as a *lack*. "Life", on the other hand, according to Pippin, signals the self-determining (or self-moving) aspect of the experiencing subject's experience of the world here.

extension. In the above quotation, Hegel supposes that the experiencing subject is aware of this, and so vacillates between judging *x* to be *F*, and realizing that since it is only because of his own whim that *x* is *F*, judging that it could just as easily be that *x* is not *F*. Thus, the shape that the subject splits-up the world into is the shape that the world has, and as the subject sees this he dissolves this shape, only to form another, which is then dissolved in turn.

Now *this* sounds very much like idealism, but we must be careful to see just what kind of idealism it is, and to understand that it is *not* a position that Hegel endorses. Hegel is not here presenting an idealist á la *Berkeley* (before God's mind enters the picture, that is).¹²⁸ It is not the case that for Desire, Life consists in its sense-impressions or that in changing from *F* to not-*F* anything about the object itself changes. The world *exists* independently of how Desire conceives it.¹²⁹ This much, Hegel takes to already have been established by Kant in the first *Critique*. What are determined by Desire are just the content of the conceptual rules under which this existence is subsumed. The object, in a sense, stays as it is, untouched. What changes is the meaning of the term 'F', and so the object's *status* as an *F*. Consider a game of soccer. What happens on the field—who kicks the ball, where the ball goes, whether someone touches it with his hands, etc.—is determined by the world. Whether such activities and events count as violations of the rules of soccer, on this picture, is determined by what those rules are. Since it is Desire that determines what the *conceptual* rules are under which its object is subsumed, it is Desire that determines what counts as *F*. Hegel's idealism here is an

¹²⁸Berkeley, (1998).

¹²⁹“It is this very flux, as self-identical independence which is itself an *enduring existence*, in which, therefore, they [the differences on which Desire bases its sorting practices] are present as distinct members and parts existing on their own account”(PS §169).

idealism about what it is *to follow a rule*. The account under consideration is that following a rule consists in nothing more than certain of the activities of the spontaneous understanding. I.e., we make rules for ourselves and determine, by ourselves, what counts as following them.

There at least two problems with this picture. The first can be brought out by considering the following scenario. I go about my house touching various pieces of furniture and saying aloud each time ‘glorp’. Suppose this were a rule-governed process of *predicating* ‘glorp’ of some of the pieces of furniture in my house. Suppose further that rule-following is as the account that we are considering conceives it, so that I cannot go wrong in my judgments of the glorpness of things. Consider the following situation. I have gone all through my house (ostensibly) judging of certain things that they are glorp, and of certain others that they are not glorp. At some point, I judge of my couch that it is glorp. Later, I judge of it that it is not, and never has been, and *further*, that when I earlier judged that it was glorp, I was mistaken. We seem to run into the following problem.

- (1) The couch is glorp iff I judge that the couch is glorp.
- (2) I judge at t_1 that the couch is glorp.
- (3) The couch is glorp.
- (4) I judge at t_2 that the couch is not glorp.
- (5) The couch is not glorp.
- (6) The couch is both glorp and not glorp.¹³⁰

As if this were not troubling enough, my *recognition* at t_2 that at t_1 I judged that the couch is glorp provides the material for the following, similar argument.

- (7) I judge at t_2 that at t_1 I judged the couch to be glorp, and I judge, at t_2 , that I was, at t_1 , wrong to do so.
- (8) My judgment at t_1 that the couch is glorp is incorrect iff I judge at any time that it is incorrect.

¹³⁰We can stipulate that in asserting that the couch is both glorp and not glorp ‘glorp’ is used univocally, to say of the couch that it is both glorp and not glorp at the same time, with the same part of itself, in relation to the same thing, etc.

- (9) My judgment at t_1 that the couch is glorp is incorrect.
(10) I judge at t_1 that the couch is glorp, and I am incorrect to do so.

As (6) is a contradiction, and (10) is *ex hypothesi* impossible, something has certainly gone wrong here. This argument proceeds, however, only from the assumptions that the account under consideration is correct and that we sometimes judge differently at different times or judge ourselves to have made mistakes. Since the latter two premises are clearly true, the view under consideration—that conceptual rules are applied correctly or incorrectly solely in virtue of the judging activity of the experiencing subject—is clearly untenable. Notice that there is no room here for a move that would modify the claim to apply only to *justified*, *correct*, etc. judging. Any such move would simply import a locus of correctness other than the experiencing subject, and in doing so would undermine the account in question.

This worry, while serious and important, is not Hegel's. Hegel's worry about this picture of rule-following stems directly from the constraint that we have already seen the Transcendental Deduction impose on any acceptable account of rule-following: that it must make following a rule recognizable as such to the one who follows (or attempts to follow) it. Hegel's worry is that this condition is not met by the present account because, according to it, there is no distinction to be made between *seeming* to follow a rule, and *actually* following one. Hegel puts this point in terms of the *objectivity* of rule-following, i.e., in terms of the accordance of the object with the rule that is applied to it.

On account of the independence of the object, therefore, it [Desire] can achieve satisfaction only when the object itself effects the negation within itself; and it must carry out this negation of itself in itself, for it is *in itself* the negative, and must be *for* the other what it *is*. (PS §175)

What the experiencing subject needs is a standard for the correctness and incorrectness of his judgments that is *independent* of it, that is *not* it. It is not sufficient that Desire predicate arbitrary concepts of its object. It must be the case that this object, in some sense, is *genuinely* subsumable under these concepts. The object must *be* for the subject what it *is*. Thus, Hegel is here pointing out that, as long as Desire is the only available standard for what counts as the correct application of a conceptual rule, *for Desire*, there *is* no such standard. That is, as long as whatever Desire does is “correct”, there can be, for Desire, no distinction between correct and incorrect at all, no *getting it right*.¹³¹ Desire, thus, cannot recognize following a rule as such because to it, everything *seems* to count as following a rule, and there is no basis *other than* this seeming for it to judge.

What Desire requires is that there be a standard for when it correctly follows a rule that is available to it as such. What Hegel suggests is that so long as it is Desire that we are considering—so long as the experiencing subject is the only standard under consideration—Desire can never achieve this end. Consider again the case in which I go all through my house ostensibly predicating ‘glorp’ of certain pieces of furniture, and ‘not-glorp’ of others. There is nothing about the *objects* of my predications that speaks to whether these predications are correct or incorrect. Of course, there *would be* something about the object that would make these predications of ‘glorp’ correct or incorrect *if the content of that concept were fixed*—namely, that they are or are not glorp—but the issue at hand is how such content fixing is possible. Furthermore, there is nothing about my *past* uses of ‘glorp’ to help here either. It might be thought, that is, that my past uses of ‘glorp’ would form some sort of pattern in accordance with which I could project future

¹³¹“One would like to say: whatever is going to seem right to me is right. And that only means that here we can’t talk about ‘right’”(Wittgenstein, 1953: §258).

correct and incorrect uses of ‘glorp’. The problem, of course, is that there must also be some standard of correctness for what can be extrapolated from this pattern, and here again I am left with only my own judgment about how it *seems* to me. Thus, on the Desire model, the experiencing subject looks not to the object to determine whether he has used ‘glorp’ correctly, but only to his own judgment on the matter.

[S]elf-consciousness is thus certain of itself only by superseding this other that presents itself to self-consciousness as an independent life; self-consciousness is Desire. Certain of the nothingness of this other, it explicitly affirms that this nothingness is *for it* the truth of the other; it destroys the independent object and thereby gives itself the certainty of itself as a *true* certainty, a certainty which has become explicit for self-consciousness itself *in an objective manner*. (PS §174)

This is how Hegel describes the first phase of this process, whereby the experiencing subject turns away from objects as the standard for correctness, and towards himself. We have already seen how he describes the second phase, in which this subject sees that, as long as all he has to base his decision on is how things appear to him, there can be no possible distinction for him between correct and incorrect uses.

What the experiencing subject needs, then, is a kind of resistant force, something that is able to act as a standard for correct and incorrect uses of a concept that is *independent* of his own judgments of correctness and incorrectness. As Hegel puts it,

Consciousness has for its object one which, of its own self, posits its otherness or difference as a nothingness, and in so doing is independent. (PS §176)

The idea here is that, if the experiencing subject is to judge correctly, there must be a standard for correct judgment with which he can agree but which at the same time is independent of his own judgment. The natural first place to look for such a standard was the world, but we have seen that *for the experiencing subject* this will not do. The world

is the *object* of his judgment, but is not a standard that he can use to determine the correctness or incorrectness of his judgment about it. This is because, until the content of his concept is fixed, nothing in the world can make a judgment about that content correct or incorrect.

A helpful context in which to think about this problem is that in which it comes to the attention of an experiencing subject that he has *misused* a concept. Hegel's point is that such a mistake can only *come to his attention*—as opposed to his unilaterally *deciding* that he has made a mistake—if there is some matter of fact about what counts as a mistake that is independent of his own judgments. I.e., the experiencing subject must be *corrected* by something or someone. What Hegel sees is that for the Kantian constraint on rule-following to be met, whatever does this correcting of the experiencing subject must be recognizable to him as having the *authority* to do so. The world does not have the power to correct the subject, and so the issue of authority does not so much as arise regarding its role in this story. As we have seen, the experiencing subject himself *would* have this authority, but can never be in a position to correct himself. (For reasons that we saw earlier, he always either agrees with himself or runs into contradiction.) What Hegel then confronts is the question of what *could* have this authority. His answer is that,

Self consciousness achieves its satisfaction only in another self-consciousness. (PS §175)

Later he adds that,

Self-consciousness exists *for a self-consciousness*. Only so is it in fact self-consciousness; for only in this way does the unity of itself in its otherness become explicit for it. (PS §177)

Only by standing in some relation to another experiencing subject—which Hegel here calls “achieving satisfaction in” and “existing for”, about which we will have more to say

in a moment—does the experiencing subject have available to him a standard for correctness and incorrectness. Two questions immediately arise: First, how is Hegel’s answer supposed to work; and, second, is it successful?

How is it that an experiencing subject can come to recognize anything other than himself as having the authority to determine what is correct and incorrect vis-à-vis his conceptual (rule-governed) judgments? The proposal we are considering is that this is possible only if this other thing is another experiencing subject. This proposal has two parts, which it will be important for us to keep distinct. First, there is the claim that another experiencing subject *can* provide a standard for the correctness and incorrectness of the conceptual judgments of an experiencing subject. Second, there is the claim that *only* another experiencing subject can do this.

To see the appeal of this position, let us consider again the failures of the world and the experiencing subject to provide a standard. Consider the situation that we proposed earlier had to be possible if the experiencing subject is to act according to a rule and recognize his doing so as such: the situation in which he is *corrected*. We saw that the world could not correct the experiencing subject because the experiencing subject is never in a position to recognize the “corrections” of the world as such. Since he is making judgments *about* the world, he must always base these judgments on how the world *appears to him*. Thus, the experiencing subject in this case really uses only himself as his standard. He may *take* himself to recognize the world as authoritative, but for him to do so is a meaningless gesture. It is really only his whims that govern his behavior. (Of course, his whims are *causally* affected by the world, and while such causal relations will have *some* place in the justificatory story that will eventually have to be told here, it is not

at the level of being corrected and recognizing this as such.) Further, the experiencing subject cannot be his own standard because, as we have seen, this simply means that whatever he takes to be correct will be correct, and this is just to say that he has no standard for correctness at all.

How, then, is another able to correct the experiencing subject in such a way that the latter can recognize this correction as such? We will begin with a broad-strokes picture of how Hegel thinks this works, and then focus on particular details as the need arises. Suppose, then, for the moment that the experiencing subject can recognize another experiencing subject as such.¹³² Suppose further that the experiencing subject can recognize that another experiencing subject makes certain judgments.¹³³ The experiencing subject, then, is confronted with the following situation. He encounters a being that he takes to be like himself (another experiencing subject) and who makes judgments about the world as he does. As Hegel puts it,

Thus the movement is simply the double movement of the two self-consciousnesses. Each sees the *other* do the same as it does; each does itself what it demands of the other, and therefore also does what it does only in so far as the other does the same. (PS §183)

Suppose further that there will be times at which the experiencing subject agrees with the other in the judgments that he is willing to make, and there will be other times in which he finds himself disagreeing with the other. Hegel's idea is that since the experiencing

¹³²The conditions under which this is possible is a subject that is taken up by Hegel briefly in the *Phenomenology of Spirit* in the section about the Life and Death Struggle, and then given a fuller treatment in *Elements of the Philosophy of Right*. Hegel's thought in the former seems to be that while animals are such that they can have desires and act to preserve their own lives, humans—qua self-conscious conceptual agents—can, in addition to this, reflect on those desires and decide to give up their lives. Thus, the thought seems to run, one sign that something is a self-conscious conceptual agent is that it is capable of engaging in a life-and-death struggle. Of course, animals can *risk* their lives, and so it is unclear what distinguishes *this* from engaging in a life-and-death struggle.

¹³³We can suppose that he does so in a way along the lines explicated by Sellars' account of the meaning rubric. I.e., he classifies the other's utterances as more or less functionally analogous to his own.

subject recognizes the other as the same sort of thing as himself, he must take the other's judgments as authoritative just as he does his own. To borrow once again from Kant, part of what is involved in the use of an object-concept is the endorsement of certain counterfactual conditionals about what manifold of intuitions an experiencing subject would have if he were, say, situated elsewhere in space, or if background conditions were different etc. In a sense, when one experiencing subject encounters another (and recognizes him as such), he takes that other as a version of himself elsewhere, and by doing so he is, by the very fact of *his* judging a certain way, committed to that other's making certain corresponding judgments. If this is the case, however, then when the two disagree, something has obviously gone wrong; a contradiction of sorts arises. *This*, Hegel thinks, is exactly the situation for which we have been searching. Confronted with a contradiction, the experiencing subject and the other must work out who, if either, is right and who is wrong. In doing so, the *community* that is thereby formed provides a *standard* for correct and incorrect uses of the concept in question.¹³⁴

One concern that immediately arises from considering this picture is that it bears a striking superficial resemblance to the picture we rejected earlier of the individual setting his own standard for correctness and incorrectness.¹³⁵ On that picture, instead of having two experiencing subjects that contradict each other, we had one experiencing subject

¹³⁴Hegel's first attempt at understanding how this process works itself out—in an early version of the *Phenomenology* published in English as *Hegel and the Human Spirit*—focuses on the need for the parties to agree, and is modeled on a loving marriage that produces a baby. In the *Phenomenology* he focuses instead on the essential role that possible *disagreement* plays, and so begins instead on the relation between Master and Slave. (Each party takes up the cause of his own judgment and tries to *master* or *enslave* the other. They each see that the other can provide a standard for them, and since each wants to be *correct*, tries to force the other into agreement. Of course, Hegel sees that *this* kind of agreement actually does neither any good, and the remainder of the *Phenomenology* is devoted to ascertaining what kind of agreement will actually work.)

¹³⁵This similarity is explored at length in Blackburn, (1984).

whose judgments at different times were contradictory. There we rejected that picture because of this contradiction. Here we have just said that the contradiction between two experiencing subjects provides the key to solving our present difficulties. If this is really to be a solution, ours or Hegel's, there must be some disanalogy between the two cases. It must be that, despite appearances, the way that diachronic time-slices of a person interact with one another vis-à-vis rule-following is different from the way that different experiencing subjects do so synchronically. In what does this difference consist?

The first thing to notice in considering these two cases is that *what* is taken as the standard for correctness and incorrectness in the two cases is not analogous. In the case we considered earlier, correctness was determined by whatever the one experiencing subject considered correct at any given time. This, we saw, generated contradiction, and failed to provide any standard to the experiencing subject. The analogy in the current case would be if the standard of correctness was determined by whatever either experiencing subject considered correct at any given time. Clearly, this would generate a contradiction (between disagreeing experiencing subjects) and would give neither subject any further standard to which to appeal. This, however, is *not* the standard in this case, and seeing what is will bring us to a further, more important asymmetry.

In the case under consideration it is not the judgment of any particular experiencing subject that acts as a standard for correctness and incorrectness of a judgment. Rather it is the agreement of the *community* of experiencing subjects that does so.¹³⁶ Of course, one might now wonder whether a similar move can be made for the

¹³⁶Such agreement, it should be noted, is—if communal agreement is to serve the role that it is prescribed by the Transcendental Deduction—a *regulative ideal* of all experiencing subjects. I.e., we are stipulating all experiencing subjects necessarily have the goal of conceiving of themselves as single, unified subjects persisting through time. The means to this goal essentially involves having a standard of correctness and

individual considered as a community of time-slices. Why can't it be that if the individual agrees with himself over time, or comes to do so, that this agreement can act as a standard for each of his individual time-slices?

The answer to this question is to be found in the nature of the agreement so reached. As we noted earlier, while it may be the *agreement* of the community that provides a standard for correctness and incorrectness, this is only a possible standard because it provides for the possibility of *disagreement* of particular members of the community that is *recognizable as such* to those members. It allows for a member to acknowledge that he has been *corrected*. Consider the senses we might try to give of the time-slices of an individual correcting one another. On the one hand, it might be that a current time-slice corrects a past time-slice. This might be what happens in our previous example when I pronounce that my couch is not, and never has been glorp. The trouble here is that the past time-slice is not corrected in a way that is recognizable as such to him for the simple reason that he *exists only in the past!* No "correction" can ever reach him, so it cannot be he that is corrected by the present time-slice.¹³⁷

So consider instead whether it might be the case that a past time-slice corrects a present one.¹³⁸ In order to make this scenario different from the one in which whatever appears to the present time-slice to be correct is correct, we must have this present time-

incorrectness for object-concept-employing judgments. *If* the only way to have such a standard is to have community agreement, then all experiencing subjects must take the reaching of such agreement as a further intermediary goal, i.e., as a regulative ideal.

¹³⁷Of course, there is a sense in which a person in the past can be corrected by a person in the present. We take ourselves to have corrected Ptolemy. As we will see farther on, though, making sense of the possibility of such a correction requires a fairly sophisticated story.

¹³⁸Such is the practice that Robert Brandom takes entire communities to engage in Brandom, (2002). The problem for Brandom then is two-fold. First he has to show that such a solution avoids the difficulties to follow. Second he must show, since he is a *social*-inferentialist, why Blackburn's picture of the individual engaging in such a practice is essentially disanalogous to the community's doing so.

slice in some sense recognize the authority of the past time-slice vis-à-vis the correctness and incorrectness of his judgments. I.e., just as one individual recognizes the authority of his fellow community members over himself, so the present time-slice will analogously have to recognize the authority of his past (and future) selves over his present one. Remember that one problem we encountered with the individual experiencing subject was that even if he had a pattern of past behavior before him when determining what to do next, there was no standard available to him for determining how this pattern is to be *continued*. That is, for this solution to work it must be the case that the present time-slice corrects his current use of a concept to bring it into conformity with his past use(s). The problem is in determining in what such a rectification would consist. It seems that here we run into the same problem anew: that whatever will seem to him to be rectification will be rectification. This is not, of course, because his memory of what has happened in the past is in any way deficient, but rather because nothing about what has happened in the past determines what *ought* to happen in the present and future. That is, if it appears to the present time-slice because of his *recognition of his past patterns of behavior* that he has gone wrong, there seems to be no additional fact of the matter to his having gone wrong. Because he must interpret this pattern of behavior in one way or another, as determining that such-and-such rather than such-and-such is the correct way to proceed, it is still *his whim* (regarding what “fits” with his pattern of past use) that is determining what is correct and incorrect for him. Instead of attempting to use the passive world as a standard, he now attempts to use his passive past selves as one. Both run aground on the same difficulty.

Our task now, then, is to show how the community solution escapes this fate, how it is possible for something other than the experiencing subject's own whim to act as a standard for him for what counts as correct and incorrect. Here is the proposal. An individual experiencing subject judges that Fx . The community corrects him; they assert that $\sim Fx$. The experiencing subject can then either change his judgment to $\sim Fx$, or not. In the case in which he changes, it seems as if we have succeeded. The community has been able to act as a standard for his use of 'F' and 'x'. What about the case in which he does not change? If such a case is possible, does it not follow that the success of the first case was a mere illusion? Is it not the case that his accepting the community as his standard for correctness and incorrectness, like his using his past selves as such a standard, is subject to his whim?

The answer here is that it is not. For consider what it is for the experiencing subject to *recognize the community as the standard for correctness and incorrectness as such*. Surely this must consist in, at least for the most part, *accepting* the judgments that the community accepts, and rejecting those that it rejects. Of course, we tried to grant this much to the time-slice individual as well, and there the problem was that what counted as acceptance was determined by the individual's whim. Here, however, that is not the case. What counts as acceptance is what the *community* accepts as counting as acceptance. To stipulate that the experiencing subject recognizes the community as the standard for correctness and incorrectness is already to concede that he will do whatever the *community* determines will count as recognizing the community as the standard for correctness and incorrectness. His whim no longer comes into the picture at all. Such an experiencing subject has already *ceded the authority* to the community to determine what

counts as correct and incorrect behavior and *what counts as taking the community as a standard*.¹³⁹ (This is what Hegel means when he writes about “the positing of the will as the will of an ‘other’, and specifically of will, not as a particular, but as a universal will”(PS §230).)¹⁴⁰

An individual time-slice has no recourse to such a move, for the problem iterates there as it cannot here. His “community” of earlier time-slices can have nothing to say on the matter of what will count as following their pattern that he cannot override. He can always construe his actions as following the pattern that he has set himself. The genuine community, however, is in a position to ensure that this is not the case. They are there (or more importantly *then*) to correct him in his construal of their corrections. The members of the genuine community *act along with* the experiencing subject in a way that his past and future time slices do not. They are present to monitor, and (if necessary) change, his behavior according to what they construe as the correct pattern. *This* is the crucial difference between the time-slice “community” and the genuine one. The rule-governed member of the genuine community is genuinely *correctable* (or *uncorrectable*, but even then still wrong).

¹³⁹Of course, communities can go wrong, and it is possible for an individual to refuse to accept the community’s agreement on some particular matter of fact as authoritative. What is essential in such cases is that the individual refuses to accept *this particular* community’s agreement as authoritative, but not the judgment of some better situated community. Just how a community can go wrong, and in what better-situatedness consists are issues that we will touch on at the close of the current discussion, but which warrant a paper all to themselves.

¹⁴⁰We have now described, at least in outline, in what recognition of the community’s authority over the correctness and incorrectness of the individual experiencing subject’s judgments consists. It is important to note that this is distinct from the issue of *why* such a subject could, would, or must arrive at such recognition. The answer to this question is that the experiencing subject must have *some* standard of correctness and incorrectness for his judgments, and *he himself* will not do, etc. So the question, for instance, of whether the individual’s acceptance of the community as his standard will depend on his whims is answerable firmly as ‘No’. Given the necessity of the success of the Transcendental Deduction, if the arguments above are sound, he has *no choice but* to accept communal agreement as his standard of correctness and incorrectness.

The main point here is that granting authority to a community consists in taking that community's agreed-upon judgments as one's standard for what is correct and incorrect, and doing this in such a way that it genuinely affects one's own patterns of behavior (in accordance with the standard for such change established by the community's agreed-upon judgment). The time slices of an individual are not in a position to produce an agreed-upon judgment of any sort, and are not in a position to rule further on what counts as conformity. One can try to take one's past use as a standard, but without further supplementation, nothing counts as succeeding or failing to do so. Taking the agreement of a compresent community as a standard, on the other hand, allows for the agreement of the community *about what counts as agreement* to do so. A community is always *present* to hand down further verdicts. Past time slices judge as they judge and do nothing else. So, a compresent community is in a position to act as a standard for correctness and incorrectness, while time slices are not.

At this point, I would like to depart more significantly than I already have from specifically Hegelian concerns in order to address a particular objection that might arise to the proposal at hand: that the agreement of a community can act as a standard of correctness and incorrectness for an individual concept-user. Just as we were able earlier to construct arguments showing the incoherence of the individual experiencing subject taking his own judgments as the standard of what is correct and incorrect, so we seem able now to construct an analogous argument about the judging activity of the community. Remember that there our example was of the experiencing subject going around his house judging of his various pieces of furniture that they were either glorp or not glorp. Since his own judging activity was being taken as the standard, whatever he

called *glorp* was glorp, and whatever he called *not glorp* was not glorp. Here, since we are taking it that the standard for correctness and incorrectness is the community's agreed-upon judgments, we can construct the following argument.

- (1) The couch is glorp iff the community agrees that the couch is glorp.
- (2) The community agrees at t_1 that the couch is glorp.
- (3) The couch is glorp.
- (4) The community agrees at t_2 that the couch is not (and never was) glorp.
- (5) The couch is not glorp.
- (6) The couch is both glorp and not glorp.¹⁴¹

This seems to show that the community's being the standard of correctness and incorrectness coupled with the fact that the community at different times agrees to contradictory claims shows that using the community as such a standard is incoherent.

Again, the following argument has an even more disturbing conclusion.

- (7) The community agrees at t_2 that at t_1 the community agreed that the couch is glorp, and the community agrees at t_2 that it was, at t_1 , wrong to do so.
- (8) The community's agreement at t_1 that the couch is glorp is incorrect iff the community agrees at any time that it is incorrect.
- (9) The community's agreement at t_1 that the couch is glorp is incorrect.
- (10) The community agrees at t_1 that the couch is glorp, and is incorrect to do so.

This, of course shows, using nothing but what the community has itself agreed, that either the community can go wrong, which is ruled out *ex hypothesi*, or that an individual's using the community as a standard of correctness and incorrectness is as incoherent as using himself would be. If it is true that the only way to provide an individual experiencing subject with a standard for correctness and incorrectness is via the community, and that doing this is a necessary goal of any experiencing subject, then there must be some response available to these two arguments. Furthermore, as we took the

¹⁴¹ Again stipulating that the community agrees that the couch is both glorp and not glorp at the same time, with the same part of itself, in relation to the same thing, etc.

analogues of these two arguments to be reasons for rejecting the individual's potential to act as his own standard of correctness and incorrectness, to continue to do so requires that the response we give to these arguments be one that is not available for transposition back to that case.

Before we delve into the solution to these difficulties it will be worth noting that while the above argument is a rudimentary example of intra-communal contradiction, there is a more interesting kind of contradiction closely related to it. This is the case in which the community licenses some material inference about the manifold of intuitions that is proven fallacious, i.e., when the community makes a *prediction* that is *disappointed*. Notice that, using only the resources thus far provided, we are forced to count this as a kind of intra-communal contradiction. The community agrees to both the judgment that p (at the time at which the prediction is made), and the judgment that $\sim p$ (at the time at which the community judges that the prediction had failed). We should like to say that the community is in the position of both *having endorsed* a prediction and *now accepting* its falsification, but that it is never in the position of explicitly endorsing *both*. As the matter currently stands, however, we do not have available to us a way of accounting for such a *change*. What we thus have is a single community whose two (temporal) parts each endorse a different judgment. Without a way to further individuate communities, such a situation must count as a community's endorsing a contradiction. Of course, it is the goal of what follows to remedy this situation.

In considering this pair of arguments, and the more sophisticated relative of them discussed just above, we must be very careful in what *kind* of conclusion that we draw from them. In particular, there is reason to think that the conclusions we have drawn

above, as presented, are too strong. What we have in each of these arguments are cases in which the community has run into self-contradiction. It agrees both to one judgment and to its negation. The conclusion we drew above was that the community was not an adequate standard for the correctness and incorrectness of judgments. What is certainly right to conclude here is that any community *that runs into this kind of difficulty* is inadequate to that task. We assumed above that it is simply a *fact* that communities do so, and so it is simply a fact that communities are inadequate to providing a standard for correctness and incorrectness to the individual. We must now investigate this purported fact more closely.

Suppose for a moment that *some* community ran into a contradiction of the sort depicted in the above two arguments. Suppose further that another community did not. The proper conclusion would seem to be that the former would be inadequate to the task of acting as a standard for correctness and incorrectness, but the latter would do just fine vis-à-vis these arguments. Now, such a situation is possible only if it is possible for there to *be* two communities. For instance, if by ‘community’ we just mean all of the experiencing subjects in the world, then clearly there cannot be *two* such communities, and any contradiction found in this community’s judgments infects the entire idea of using the community as a standard. The situation described here, however, of *two* communities, one of which runs into such a contradiction and one of which does not, seems entirely possible. The question, then, is whether there is any compelling reason to think that there can be only a single community of this sort, or whether there is a more fine-grained distinction that can be made here.¹⁴²

¹⁴²The motivation for this move is exactly parallel to that for the position we discussed earlier that attempted to remedy the difficulties for the *individual* by dividing him into time-slices. Thus it will be

To answer this question, consider for a moment the specific role that we have been assigning to the community. What we needed was a standard for the correctness and incorrectness of the application of conceptual rules to manifolds of intuitions. One plausible way of thinking of a community, then, is just as that which has the function of providing such a standard (via the agreement of compresent independent experiencing subjects). Things with functions, however, can succeed or fail in performing their functions, and as we have now seen, the community is no exception. A community can fail to provide a coherent conceptual scheme to its members whenever it runs into contradiction. As it is the *goal* of the members of the community, however, to *have* a coherent conceptual scheme, such successes and failures of the community must be recognizable to its members. As we have emphasized throughout, goal-directedness requires that the ends that one pursues and the means that one takes be recognizable as such. This, in turn, requires that the goal-directed agents be able to make a distinction between merely seeming to recognize a goal or means as such, and actually doing so. This further requires that the ability to be *corrected*. What all of this suggests is that there must be some way for the members of a community that fails to fulfill its function, to be corrected. What is puzzling here, though, is that if the community agrees to a contradiction, and the standard of correctness and incorrectness of the judgments of the members of that community is provided by the agreement of the community, then it would seem that the members of the community cannot, in fact, be corrected. This is because the standard of correctness *is the very thing about which they would need to be corrected*. So either the task we have outlined is impossible, or there must be some

essential to keep track of the *differences* between the strategies employed here to make ‘community’ more fine-grained, and those employed there to make ‘individual’ so.

further standard of correctness and incorrectness that comes into play in cases of intra-communal contradiction.

We now have *two* distinct puzzles to consider. The first is how we individuate communities in a non-arbitrary way more fine-grained than merely the collection of all experiencing subjects. The second, which has just emerged in the course of our argument, is how it is possible for the members of a community to be corrected *by something other than the community*, and to recognize this correction as such. We can say a bit about both puzzles here, although the latter is a substantial issue in need of its own, independent study.

First, then, we will address the topic of how to individuate communities. Once again, it is worth our while to look to Hegel's *Phenomenology* for assistance. In particular, in the closing sections of his chapter, "Spirit", Hegel considers a situation in which two communities (Spirits) confront each other, each maintaining a consistent set of standards of correctness and incorrectness (Notions) for their members. While this particular situation does not concern us here, how Hegel individuates these communities does. He writes,

Each of these two self-certain Spirits has no other purpose than its own pure self, and no other reality and existence than just this pure self. But yet they are different; and the difference is absolute because it is set in this element of the pure Notion. (PS §671)

The two communities (Spirits) differ just insofar as their conceptual schemes (Notions) differ. That is, these communities are distinct insofar as the *rules of inference*. The suggestion we can then garner from Hegel is that we should individuate communities by *conceptual scheme*.

Of course, if we take this suggestion, we run the risk of eliminating all too easily the contradictions that we have been considering. That is, one might worry that, if we individuate conceptual schemes with too fine a grain, then those who agree to p will constitute one conceptual scheme, those who accept $\sim p$ will constitute another, and contradiction will be conceptually impossible. This is a real worry that, especially when combined with issues surrounding incommensurability, must be addressed. It is not, however, directly relevant to the current discussion. Remember that the community we are considering is one that *itself endorses both* of two contradictory propositions. The entire community agrees, we can suppose, that both p and $\sim p$, and so cannot be fractioned off into those that agree that p and those that agree that $\sim p$. As to why any such community would knowingly do such a thing, we have already seen the answer. This is roughly the situation in which a community finds itself when it endorses a material inference—an inference about the way the manifold of intuitions ought to be—that is later shown to be fallacious. I.e., it is the position a community is in when it makes a prediction that fails. It is in the position of both endorsing the prediction and its falsification.

Describing this process as one in which a community fails to fulfill its function, leads us to the second of our puzzles: how it is possible for the members of the community to be corrected by something other than the community, and to recognize this correction as such, for we have been arguing that for a person or group of people to have a goal they must be able to represent the conditions of attaining or failing to attain that goal, and that doing this requires that they be subject to correction that they can recognize as such. Up to this point we have been working through the suggestion that it is

consonance of communal judgments that provides a standard of conceptual correctness and incorrectness for the individual experiencing subject. The suggestion on the table now is that the community too must be held to some standard, not for correct and incorrect applications of concepts—we have already conceded that, in this respect, whatever the community says, goes—but rather for its success in providing such a standard to the individual. The puzzling part of all of this is that it would seem that our arguments regarding what the individual might use as a standard seem to lead to the following conclusion: nothing but agreement with a community can act as a standard. This would seem to show that the members of the community cannot appeal to *anything* for their standard for the success of the community itself. The question, then, is how it is so much as possible to regard a community as correct or incorrect in its endorsement of a particular purported conceptual scheme.¹⁴³

Given our answer to our first question, however, about individuating communities, a solution to this puzzle seems well at hand. Suppose that we can individuate communities by conceptual scheme. Then just as a community acts as the standard of correctness and incorrectness of the judgments of its members, it certainly seems possible that a community *of communities*¹⁴⁴ can act as the standard of correctness and incorrectness of the judgments of the members of a community vis-à-vis their success and failure in providing themselves with a conceptual scheme. That is, the

¹⁴³Wright, (1981) argues that nothing can act as such a standard. Brandom, (2002) argues that communities can use other communities as standards in the way that common-law judges use others common-law judges as standards. Rosenberg, (1980) argues that communities can use other communities as standards by employing a Sellarsian criterion for justified theory change.

¹⁴⁴Perhaps the most relevant example of such a community of communities is what we call the *scientific* community. The proponents of competing theories share a conceptual scheme, and the purpose of the scientific community is to adjudicate among them. This is certainly a model that would be congenial to Kant and Hegel.

solution to our previous problem of what can provide a standard of correctness and incorrectness for an individual will have an analogue one level higher. The agreement of a community of communities about whether a particular community has succeeded or failed in its goal of providing a conceptual scheme for its members can act as a standard of success or failure for an individual community.

It is important to notice the differences between this suggestion and the suggestion we considered earlier regarding considering an individual as a collection of time-slices. Both were a way to avoid the difficulties presented by de facto contradictions—the former for the community, the latter for the individual. We have argued that the proposal in the case of the individual does not work, and it is a point in favor of the current proposal that that it is not subject to the same objections that led us to that conclusion. The trouble with using the agreement of time-slices of individuals as the standard for correctness and incorrectness is that because each time slice necessarily exists at a time when no others do, it is difficult to see how any correction of one time-slice by another could possibly be made. An individual is always free in the present to take his past time-slices' behavior as consistent with his current behavior or not as he sees fit, and is equally uncorrectable by those time-slices of himself that have not yet come into existence. Such difficulties are a direct byproduct of the proposal that an individual be divided by across *time*. The current proposal, however, faces no such restriction. Individuating communities by conceptual scheme leaves open the possibility of two such communities existing concurrently, and so being in the position to correct one another.

Of course, we do not have the time to delve into the details of this suggestion here. What we do have time for is to outline a response to a fairly straightforward objection to it which will hopefully offer some insight into how a further articulation of the suggestion might go. The objection is simply this. Just as we took contradictions at the levels of the individual and the agreement of a community as reason to abandon employing these as standards of correctness and incorrectness, we can also imagine that a contradiction could arise at the level of the community of communities. By parity of reasoning this would seem to preclude using the agreement of such a community as a standard.

The beginning of an answer to this object can be found by noticing that both Kant and Hegel argue that one always needs to employ *some* conceptual scheme. Thus, in cases in which a particular conceptual scheme *fails*, the solution can never be to simply abandon that scheme, but must always also involve *replacing* it with some other one. Thus, the agreement of a community of communities is not only about the failure of certain conceptual schemes, but also about the suitability of adopting certain successor schemes. This brings us to the thorny issue of the *progress* of science. What we need, then, to respond to this objection is an account of such progress that provides a standard for rational theory change that a community in the grips of such a change can recognize as such.

One suggestion that seems amenable here is that there might be some single criterion for when one conceptual scheme is suited to replace another (such as Sellars' suggestion that the successor theory be able to explain both the successes and the failures

of the predecessor scheme).¹⁴⁵ The agreement of a community of communities that then acts as a standard for particular communities is the agreement reached at the hypothetical end of a process of successively replacing one conceptual scheme by another until an adequate one is found. As Hegel puts it in a slightly different context,

The realm of Spirits which is formed in this way in the outer world constitutes a succession in Time in which one Spirit relieved another of its charge and each took over the empire of the world from its predecessor. (PS §808)

Of course, since this agreement is hypothetical, and if real, in the far distant future, it cannot act as a standard in the same *way* that a community acts as a standard to an individual. This final community of communities is not present to correct today's communities. Rather, achieving such an agreement must act as a *regulative ideal* of such communities, pursued *by* employing the single criterion for theory change. If there were such a criterion, one which could produce univocal verdicts about when it is appropriate to replace one conceptual scheme with another, the employment of such a criterion would allow communities of communities to work towards the common goal of reaching eventual agreement with one another about a single conceptual scheme adequate for representing the world. The successor community, by providing a conceptual scheme that meets the appropriate criteria would act as an immediate check on the predecessor community, and would do so in the service of using the hypothetical end-community as a more mediate check.

The picture we would have if such a line is tenable would be as follows. The standard for the correctness and incorrectness of the judgments of a conceptual agent is the agreement of the community of which he is a part. Communities are individuated by

¹⁴⁵Pace Kuhn, (1977), Sellars, (1961). Rosenberg, (1980) argues that this criterion can be grounded in a Kantian account of theoretical *objects*.

their conceptual schemes—by the rules of inference they license, prohibit, require, etc. Such conceptual schemes, in part because they are attempts to represent *the world*, can run into contradictions. When this happens, and when there is a conceptual scheme available that meets a certain criterion of adequacy, a new community is formed that employs this successor scheme. Communities engage in and continue this process in the hopes of reconciling their own conceptual scheme with that of a hypothetical community whose scheme is ultimately adequate (runs into no contradictions). This story is, I hope, at least faintly recognizable as resembling the Hegelian notion of the necessary historical development of the absolute, and is, I think, not as implausible as that notion has been taken to be.

Chapter Six

A Kantian Account of Rational Theory Change

What we confronted at the close of our last chapter was a situation in which the community—which is the vehicle by which its members are provided with a standard of correctness and incorrectness for their judgments—itself ran into contradiction. The lesson we took away from this consideration is that any community that does this is not one that is suited to playing the role that we have outlined for it. I.e., it is not suited to provide a standard of correctness and incorrectness for the individual's judgments. The suggestion we then considered is that such contradictions need not implicate communities *in general* as inadequate to this task, if only there were a way to isolate the contradiction within the practice of a single community. Our proposal was to do this by individuating communities according to the very attempts at standards of correctness and incorrectness that they deploy for their members, i.e., by the judgments and inferences that they license, i.e., by conceptual scheme.

We are now interested in two further questions. The first is whether there are any constraints on what a community ought to do in the face of a failure of their conceptual scheme. I.e., given that they must adopt some new conceptual scheme (on pain of failing to meet their goal of providing a standard of correctness and incorrectness for their members), we are concerned to see what if any methods there ought to be for doing so.

Secondly, we must be concerned with how, if there is any such constraint, this could be possible. That is, we have argued in the preceding chapter that the only way for an individual to meet a standard of correctness and incorrectness (there for his judgments, here for the procedure of choosing a new conceptual scheme) is if there is a community in place to correct him if he goes wrong. It would seem, though, that the analogous standard is unavailable to our community. If this is the case, then the constraints on change in conceptual scheme that we have just been envisioning would be in-principle impossible.

My hope in this chapter is that by beginning with the first of these questions, we can gain some insight into what an answer to the second might be. That is, we will begin with the question of whether there are any constraints on change of conceptual scheme in the face of anomaly, and only after having answered this will begin our investigation into how such constraints are possible (inter-communally).

What we saw at the close of the last chapter was that, taking into account the Kantian origins of our rule-following considerations, there are more and less interesting ways in which a community might fall into contradiction. A less interesting way is that in which a community simply applies a concept, for no particular reason, to an object to which it had previously applied a concept incompatible with the first. What if any constraints on what conceptual scheme should replace one that has failed in this way is a question to which we will return, but it would not be unreasonable to expect that with a failure of this sort, there is no constraint other than consistency in place.

A more interesting way for a community to run into contradiction is for the community to agree that a concept has certain material implications, to institute the

inferences that constitute that concept, and for it to turn out that the inferences so licensed lead to expectations about the manifold of intuitions that are not fulfilled. That is, it is possible for a community to contradict itself by making a *prediction* about the world that turns out to be false. This is an especially interesting case given that the heart of our case for inferentialism is the Transcendental Deduction—Kant’s argument regarding the necessary preconditions for conceiving of oneself as a single, unified subject of experience persisting through time—and it will be worth our while to briefly revisit that story to see why. The thought here is that we saw in earlier chapters that the Transcendental Deduction had a good deal to say about *object*-concepts and the essential role they play in our cognition. Since we are now considering cases in which the community enforces certain object-concept-employing norms which *fail*, perhaps a further investigation into object-concepts can give us some guidance regarding in what such a failure consists, and how a community ought to (or must) deal with such failures as they arise.

Remember, then, that Kant’s story begins with the question of how it is possible to conceive of oneself as a single, unified subject of experience persisting through time. His answer is that this would be possible if one could *unite* a diachronic manifold of intuitions in a single cognition. That is, if we can think an entire manifold of intuitions in a single thought, then it must be one and the same subject of experience that *has* all of those intuitions. The question we confronted in our third chapter was how such a cognition is possible. One important part of that answer was that it is only possible if the object-concepts employed in that single cognition are inferentially articulated. To represent oneself as a single, unified subject of experience persisting through time, one

must represent one's intuitions as being necessarily connected to one another. One does *this*, on Kant's view, by placing these intuitions in *inferential* relations to one another. So, the necessary connection between, say, the tail of an elephant and its trunk, is represented by the material inference license from 'Here is an elephant's tail' to 'There will be its trunk'.

What we did not focus on in our earlier chapter, but which is of crucial importance now, is that in placing these intuitions into such inferential relations with one another one represents the necessary connections among the manifold of intuitions by, in essence, building a *picture* of way the world is using the possible judgments one can make about it. Intuitions are singular representations of determinate individuals. Concepts are the rules for placing such representations into inferential relations with one another. The use of such concepts can be better or worse, in the way that we have just seen. They can place these representations into more or less accurate relations. As the accuracy of these relations is a matter of licensing all and only those relations which lead to the matching of expected intuitions and actual intuitions, the system of relations as whole can be thought of as a kind of picture, or map.

What is crucial to our present purposes is that this map is a map *of* something: of the single, lawfully governed world of objects existing in space and time. Vis-à-vis the Transcendental Deduction, representing this world is *the* purpose of employing concepts. Kant's thesis is that it is *by* representing the world as such that we come to represent ourselves as single, unified selves persisting through time. The world and the self come together in one package, for Kant, and they do so by jointly making it possible for us to conceive of ourselves as *experiencing a world*. To take one's experiences as experiences

of a world, is only possible by taking those experiences to be those of a single self. To take one's experiences to be those of a single self is only possible by taking them to be experiences of a single world.

Given our present concerns, parts of this story merit emphasis. In particular, because it is inconsistencies between expected structures of the manifold of intuitions and actual structures of this manifold that concern us, and since it is object-concepts that give us our expectations regarding this structure, it will behoove us to spend some time understanding Kant's views on object-concepts and objects better than we have up to now. Speaking to just the way in which objects and object-concepts figure into the unity of the self, Kant writes,

Thus the original and necessary consciousness of the identity of oneself is at the same time a consciousness of an equally necessary unity of the synthesis of all appearances in accordance with concepts, i.e., in accordance with rules that not only make them necessarily reproducible, but also thereby determine an object for their intuition, i.e., the concept of something in which they are necessarily connected; for the mind could not possibly think of the identity of itself in the manifoldness of its representations, and indeed think this *a priori*, if it did not have before its eyes the identity of its action, which subjects all synthesis of apprehension (which is empirical) to a transcendental unity, and first makes possible their connection in accordance with *a priori* rules. (A108)

Here Kant identifies the concept of an object with the concept of "something in which the manifold of intuitions is necessarily connected", and tells us that it is only by synthesizing this manifold in accordance with such a concept can we unite it in the way necessary to conceiving ourselves as single, unified subjects of experience persisting through time. Thus, the concept of an object is the concept of *that which* provides for the necessary connections between our intuitions, i.e., the structure of the manifold of intuitions. In a sense we already knew a good deal of this. We knew that object-concepts are meant to provide rules for uniting the manifold of intuitions, and that this could only

be done by supposing that there are necessary connections within that manifold. The important addition that this passage makes to our understanding of objects and object-concepts is that, according to Kant, this work is done by positing that there is a *thing* that *explains* these why the manifold of intuitions is as it is. The concept of an object is the concept of that which explains the manifold of intuitions (its existence and its structure).

Kant's thought is this. In order for us to conceive of ourselves as single, unified subjects persisting through time we must conceive the manifold of intuitions as exhibiting certain necessary connections. For instance, the necessary connections between an intuition of a short grey tail, of a big grey body, of a long grey trunk, etc. The *reason* thinking of these connections as necessary *helps* in thinking of ourselves as single, unified subjects of experience is *because* it is by doing so that we come to think of the world as single, lawfully governed world of objects existing in space and time.¹⁴⁶ For Kant, positing necessities among the manifold of intuitions and positing the existence of objects go hand in hand.

If we enquire what new character *relation to an object* confers upon our representations, what dignity they thereby acquire, we find that it results only in subjecting the representations to a rule, and so in necessitating us to connect them in some specific manner; and conversely, that only in so far as our representations are necessitated in a certain order as regards their time-relations do they acquire objective meaning. (A197/B242)

It is *by* positing certain necessary connections among the manifold of intuitions, licensing certain intuition-involving inferences, that we come to think of the world as containing objects. What makes a concept a concept *of an object* is just that it posits such necessities. So, employing such concepts always involves positing such objects. The way that we think of our manifold of intuitions as being the result of multiple encounters with the

¹⁴⁶Henceforth, the world and the self, respectively.

world—rather than, say, a mere hodge-podge collection of sensory states—is that we suppose that these intuitions are necessarily connected in various ways.

Consider again, our elephantine manifold.

- t₁: This short grey tail.
- t₂: This big grey body.
- t₃: This big, flat grey ear.
- t₄: This long grey trunk.

What distinguishes our thinking of this manifold as a mere series of images from our thinking of it as being the result of an encounter with an *object* (an elephant) is just our employing the object-concept ‘elephant’ in a judgment about this manifold. We already know that employing that concept involves thinking of these intuitions as necessarily connected in certain ways (by placing these intuitions into certain inferential relations). What Kant is proposing, then, is that thinking of our intuitions as necessarily connected just *is* to think of them as being the result of encounters with objects.

It is a key piece of the Transcendental Deduction to see that our manifolds of intuitions must be united. To do this we must posit certain necessary connections among these manifolds. To do this is to suppose that these manifolds are the result of encounters with objects. The kind of object that we posit will vary with the necessary connections we posit among the manifold of intuitions. Different necessary connections, different objects. What all of this amounts to, then, is a picture of objects as a kind of *theoretical entity*. Our concept of an object, on this picture, is a concept of that which *explains* the manifold of intuitions. Our use of object-concepts is first and foremost in the service of placing the manifold of intuitions into inferential relations with one another so as to form a coherent picture of the single, lawfully governed world of objects existing in space and time.

To posit certain necessary connections among the manifold of intuitions is to posit a certain kind of *object*, and so to posit these-and-those necessities is to posit such-and-such an object. Thus, if what we are presently concerned with is what to do in the face of a disappointment of our expectations vis-à-vis the necessary connections among our intuitions, the question that Kant would have us ask is in what sense the object that we postulated as explaining the manifold of intuitions is not up to that task. That is, Kant will construe the breakdown with which we have been concerned as stemming from the postulation of the wrong kind of *theoretical entity*. What has run our community into contradiction in this case is supposing that the wrong kind of objects exist.

We are now in a position to reframe our original question. That question was whether there are any constraints on what a community ought to do in the face of a failure of their conceptual scheme.¹⁴⁷ We can now put this question, at least in cases where contradiction arises from some failed prediction, as whether there are any constraints on what new *kinds of objects* a successor conceptual scheme can postulate in the face of the failure of its predecessor to posit the right kinds. Given the situation we have been describing, there are a number of such constraints that it would seem plausible to propose. Firstly, as the problem with the predecessor conceptual scheme is just that it has run into contradiction, the first desiderata for any successor scheme should be that it does not itself, at least *prime facie*, run into any contradictions.¹⁴⁸

¹⁴⁷For present purposes we are counting any failed prediction as a failure in conceptual scheme, and any change in the face of such failure as a change in conceptual scheme. This is because we are individuating conceptual schemes by the sum of the inferences licensed by a given community. There is, of course, reason in other contexts to speak of conceptual schemes more loosely.

¹⁴⁸Individuating conceptual schemes holistically as we are, any conceptual scheme that is proposed but in need of “tweaks”. Cf. Kuhn, (1970); Lakatos, (1970) will count as a failed scheme, which is then replaced by a new one.

For our second criteria, notice that what the predecessor scheme will have done is noticed some regularities amongst the manifold of intuitions, taken some of these regularities as necessary, and thereby posited the objects meant to explain these regularities. Now, what happens when certain predictions of this conceptual scheme go wrong is that it is shown that it was a mistake to take certain regularities to be *necessary*. The regularities, however, might very well remain, and with the abandonment of the predecessor scheme, without sufficient provisions preventing this, would be left entirely unexplained. Furthermore, a new phenomenon would also be left unexplained, namely, why it is that positing a certain kind of necessity and a certain kind of object to explain this necessity worked as well as it did for as long as it did. I.e., presumably the predecessor scheme had its successes, and these successes will have thus become part of the data in need of explanation by the successor scheme. They have become yet another regularity—the correlation of the subset of the manifold consisting of the behavior of the users of the predecessor scheme and the subset of the manifold that they sought to explain—that is in need of explanation. Thus, the second desiderata of any successor conceptual scheme is that it be able to account for the success, to the extent that it was successful, of its predecessor.¹⁴⁹ It must explain, that is, why understanding the manifold of intuitions as the result of encounters with the kind of object that the predecessor scheme postulated worked as well as it did. It will do this by explaining what features of the predecessor scheme's postulated objects connected with what features of the successor scheme's postulated object, i.e., which purported necessary connections between intuitions of the predecessor theory were real rather than mere appearance.

¹⁴⁹Cf. Sellars, (1963); Rosenberg, (1980).

Finally, as the object postulated by the predecessor scheme turns out *not* to be capable of explaining the necessary connections among the manifold of intuitions, the successor theory ought to be able to explain why *this* is as well.¹⁵⁰ That is, the successor theory ought to be able to explain the reason for the breakdown of the predecessor scheme in light of the entities that each scheme postulates.

We argued in our last chapter that the because having a standard of correctness and incorrectness was a means by which the experiencing subject could achieve his goal of conceiving of himself as a single, unified subject of experience persisting through time, the taking of this means had to be recognizable to him as such. The thinking there was that because the experiencing subject, as a result of the practical-argumentative structure of the Transcendental Deduction—had to form the *intention* to take this means, he also had to be in a position to represent to himself in what taking it would consist. The same logic should now apply to the community vis-à-vis our claim that it has as its goal providing its members with a conceptual scheme (or with a conceptual scheme that is consistent in the ways we have been discussing thus far). As the community forms the *intention* of achieving this goal, it must be the case that the community is able to represent to itself in what this achievement would consist.

The trouble will this is that we have been arguing up to this point that representation is an essentially normative affair, that it requires that there be some standard in place for what counts as correct and incorrect representing. In the case at hand, however, this would require that there be some standard in place for the community's correctly or incorrectly representing itself as having achieved the goal of providing its members with an adequate conceptual scheme. Herein lies the rub. We are

¹⁵⁰Cf. Sellars, (1963); Rosenberg, (1980).

now forced to wonder what such a standard could be. In the case of the individual's representings, we supposed that the standard for the correctness and incorrectness of these could be the community. For the community, however, this answer simply will not do, for as we have seen nothing can be its own standard of correctness and incorrectness.

How, then, we must wonder, can it be so much as possible for the community to recognize its activities as being aimed at providing its members with an adequate conceptual scheme. As we now stand in our dialectic we seem to have good theoretical reason to suppose that the community does have providing its members with an adequate conceptual scheme as a goal, and we also seem to have good *prime facie* reason to think that having a goal, and representing more generally, is something in which communities can engage. Perhaps, then, the time has come to wonder how such a goal might be concretely manifested. At this point, it is also worth remembering that in our last chapter, we took it as a necessary condition on an individual's having a standard of correctness and incorrectness that they could be *corrected*. It seems appropriate to ask, then, in what might a correction of a community's belief that they have provided an adequate conceptual scheme to their members consist.

This, however, seems to be a question to which we have a ready answer. A community takes itself to be *corrected* in its belief that it has provided a conceptual scheme to its members when what it took to be the conceptual scheme provided is shown to *fail*, is shown to lead to contradiction. More particularly, a community takes itself to be corrected when the predictions it makes regarding what manifold of intuitions will be given to its members are wrong. A community can recognize that it has incorrectly

represented itself as providing its members with a conceptual scheme just in case that conceptual scheme can be shown to be inconsistent.¹⁵¹

This, however, seems an impossible answer. Remember that in our previous chapter we considered whether an individual could use the world as his standard of correctness and incorrectness. We there argued that he could not because the world would not be recognizable to him as an authoritative corrective. The idea was that he could “predicate” ‘*F*’ of one thing, ‘*G*’ of another, etc., and nothing about the world could possibly tell him that he was right or wrong to do as he pleased. What we now seem to be saying, however, is that when the community is put in the same situation, it *can* use the responses of the world as its standard. Why, we must then wonder, is this possible for the community but not for the individual?

The answer to this question *must* be that the community does not, in fact, use the world (including among worldly things the inconsistency of its attempt at a conceptual scheme) as its standard at all. What again seems like a very straightforward case of simply “seeing” that one is mistaken must have an underlying logic that temporarily eludes us. What we were tempted to say—that a community can represent itself as having succeeded or failed in providing its members with a conceptual scheme simply by checking that conceptual scheme for inconsistency—cannot possibly, on pain of violating all the conditions we have hitherto set on representation, be the entire story.

¹⁵¹There may, in fact, be further reasons for taking a conceptual scheme to have failed, apart from just its inconsistency. Explanations can be bad because they are not unified, because they are not powerful, because they are not simple, etc. These would all be reasons to be dissatisfied with a conceptual scheme, and to replace such a conceptual scheme with one that shares its faults to a lesser degree. However, such problems do not provoke a *crisis* of the kind we have been considering, where the scheme *can no longer be used at all* and *must* be replaced on pain of its members no longer being able to conceive of themselves as single, unified subjects of experience persisting through time.

Of course, as we saw in our last section this *was not* the entire story. Our story there—of rational theory change—involved *two* communities, and suddenly a solution seems apparent. What we needed in the case of the individual was some standard for correctness and incorrectness that the individual could recognize as authoritative over his own judgments. Our solution there was that the agreement of a community of such individuals would do the trick. What we need now is some standard of correctness and incorrectness that the community can recognize as authoritative over their judgments. What seems to be worth exploring now is the suggestion that a community *of such* communities could do this work. (For clarity's sake, we can use 'coalition' in place of 'community of communities'.)

Our proposal for how such a coalition can operate as the standard of the correctness and incorrectness of the community's judgments regarding its success and failure in providing its members with a conceptual scheme will be intimately tied to the picture of rational theory change sketched above. Before we delve into the specifics of our own proposal, though, it will be worth our while to see how someone *else* attempts to solve this same problem with a solution that is, at least in spirit, similar to our own. For that reason, we will now pause to consider the proposal for communal correctness and incorrectness suggested by Robert Brandom in his paper, "Some Pragmatist Themes in Hegel's Idealism."

Robert Brandom considers the problem at hand in a slightly different context, and offers what at first appears to be an ingenious and promising solution to it. In his article, "Some Pragmatist Themes in Hegel's Idealism," Brandom is concerned, as we have been, with the necessary preconditions of the institution of conceptual norms. Also as we have,

Brandom recognizes that an individual's following such norms is only possible within the context of a community of such individuals whose agreement is taken as authoritative over correctness and incorrectness of the judgments of the individual. Finally, as we have, Brandom then wonders about the inter-communal application of such norms, whether it is possible in such cases for one community to *correct* another, and if so how this could be possible given the constraints that we have already seen that Hegel's rule-following considerations place on such a process.

In fact, Brandom puts quite a high stake on answering just this problem. As he sees it, for slightly different reasons than those that we have confronted, if this question cannot be answered, then the very idea of concepts as rules, or rule-following at all, is unintelligible. Here is Brandom.

How is it possible for an application of a concept to count as *incorrect* according to the commitments implicit in prior applications? If there is nothing to the content of the concept except what has been put into it by actual applications of it (and its relatives), how can any actual application be understood as incorrect according to that content? If it cannot, then no norm has been instituted. (Brandom, 1999: 179)

Brandom's thought is that if all one had to go on in determining whether a particular concept was correctly or incorrectly applied in a certain case were past applications of that concept, then there could be no determinate content for that concept, for *any* new application can be made to accord with all past applications given enough gerrymandering.¹⁵² Of course, we have run into such worries already at the level of the individual, and our solution there was to see that what is needed is some genuine standard, recognizable as such by the individual, for present use. That standard was

¹⁵²Pace Wright, (1981).

provided by communal agreement. The question now is what possible analog there could be at the level of the community.

Brandom gives his answer by way of an analogy. The relation of past uses of a concept by a community, present uses of that concept by a community, and future uses of that concept by yet another community are to be understood on the model of the treatment of precedents by common-law judges. The purported solution comes in two parts. The first concerns the relation of past applications of the concept to present ones.

Past applications of concepts (decisions of cases) exercise an authority over future ones. For they supply precedents that constitute the only rationales available to justify future decisions. They are the source of the content of the concepts later judges are charged with applying. [...] But reciprocally, later applications of concepts by the judges who inherit the tradition exercise an authority over the earlier ones. For the significance of the authority of the tradition, what conceptual content exactly it is taken to have instituted, is decided by the judges currently making decisions. (Brandom, 1999: 180)

Here Brandom draws our attention to the interplay of the applications of concepts by communities in the past and the applications by present ones. Present communities must take some account of communities of the past—they must either make their judgments accord with the judgments of these past communities, or declare that these past applications were, in fact, incorrect. Similarly, past communities are held responsible to present ones—their judgments are either acknowledged as precedent by present communities, or are dismissed as incorrect by them. Furthermore, the *grounds* on which one community rejects the judgments of another must be *shared* grounds. For example, the present community can reject a past community's judgment only on the basis of some *reason*, which reason is one that the past community would recognize as relevant. On Brandom's account, then, there is a reciprocal relation between past and present

communities that goes some ways towards providing a standard for correctness and incorrectness for both.

Of course, if this were the entire story, it would be hard to see—for reasons analogous to those that we considered in our last chapter vis-à-vis time slices of an individual acting as standards for the correctness and incorrectness of their own judgments—just how either such community is *really* held accountable to such standards. Isn't the present community in a position to accept and reject whatever past applications that it wishes? Doesn't this amount to having no standard at all to which it can appeal? Etc. Brandom sees this difficulty and answers as follows.

The current judge is held accountable to the tradition she inherits by the judges yet to come. [...] If they take her case to have been *misdecided*, given *their* reading of the tradition she inherited, then the current judge's decision has no authority at all. The authority of the past over the present is administered on its behalf by the future. (Brandom, 1999: 181)

Brandom's answer is that the present community is *not* the sole arbiter of correctness and incorrectness because future communities will later be in a position to judge of the present community that *it is incorrect*. Of course, it is not that any given future community is the sole arbiter of correctness either. Brandom supposes that this process is one that carries on indefinitely, and so each community is provided with a standard of correctness and incorrectness by the communities that will come after it.

The first thing to notice about this proposal is its similarity to the proposal considered in the last chapter that time-slices of an individual are strictly analogous to communities and so can provide a standard of correctness and incorrectness for a present individual in exactly the same way that a community can. Here we replace the individual with the community, and add future time-slices to the group, but little else has changed.

In fact, it seems clear that if Brandom's proposal for an inter-communal standard of correctness and incorrectness is tenable, then an exactly analogous situation ought to be able to provide an individual with a standard of correctness and incorrectness using his past and future time-slices. The idea would be this. An individual is beholden to his past time-slices because he must either make his judgments accord with these, or dismiss them as being in error. This individual's present time-slice, however, is not the sole arbiter of correctness and incorrectness because his future time-slices have the same authority over him as he does over his past. Thus Brandom's solution to the problem of inter-communal standards of correctness seems also to be a way of avoiding that problem entirely. Were his proposal tenable, it would give us a way to avoid appealing to communal standards of correctness and incorrectness altogether. The individual could do this all on his own.

Of course this is not an argument *against* Brandom's proposal. If such a way exists for the individual to act as their own standard, so be it. The task of articulating conceptual content would be that much simpler. Unfortunately, Brandom's proposal runs afoul of the same worry that we raised in our last chapter regarding the analogous proposal for the individual. The standard of the correctness and incorrectness of its judgments that past and future communities would purportedly provide the present community on this proposal would be utterly transcendent to the present community. That is, such a standard would not provide the present community with *any* standard that would be *recognizable as such*. As Brandom rightly notices, past communities alone are of no help because the present community is free to interpret them as being correct and incorrect as it pleases. It is also free to interpret its own judging activity as being an "extension" of the judging activities of past communities as it pleases. Past communities

are incapable of *correcting* present communities in any but the most superficial way. They put virtually no constraint on what the present community can and cannot permissibly judge (especially when it is taken into account that the present community is free to dismiss the judgments of past communities as incorrect).

This is the point at which Brandom's proposal about future communities is supposed to help. It is supposed to be the case that present communities are correctable by future ones, just as past communities are correctable by present ones. This, however, is of little help if one thinks, as we have been supposing, that to be correctable involves recognizing that correction as such, or even if one supposes merely that to be correctable involves changing one's behavior in response to correction. Neither of these conditions can be met on Brandom's proposal for the simple reason that the community doing the correction *does not yet exist* at the time that the community being corrected does. Future communities provide no standard of correctness *for present communities*, firstly, because the judgments of future communities are not one's that can be in any sense *known* by present communities. This means that as far as present communities are concerned what future communities will agree to is entirely up for grabs. One would suppose that future communities would agree to whatever is right, and what *appears* right to the present community is what they will judge, so the best guess as to what future communities will agree to is just whatever present communities happen to agree to now. Thus, to the present community, the future community provides *no standard at all apart from their own best judgment*. Secondly, it is not as if, if the future community *does* disagree with the past community, that anything about the past community can be changed. By the time that the future community comes to disagree with the past community, the past

community *no longer exists*. So, they certainly cannot be *corrected*, at that point, by anyone.

Brandom here would be quick to point out that it is not strictly speaking true that *nothing* about the present can be changed by the judgments of future communities. The *status* of those judgments can certainly change, and since it is their status as correct and incorrect with which we are concerned, that is enough. The idea here would be that while it is true that future communities cannot, so to speak, reach back to the present to change the *behavior* of the present communities, they can still in a sense *correct* these communities. That is, they can make the judgments of these communities correct or incorrect *by* agreeing or disagreeing with them.

Again, however, the problem here is that while it may be true that the judgments of the present community can be *made true* by the judgments of the future community, more is needed for present purposes. What we are in search of is a standard of correctness and incorrectness *that is recognizable as such to the communities in question*. Remember that it is the *goal* of such communities to provide a standard of correctness to their members, and as such, doing so must be recognizable to them. There must be a standard against which they can judge their success and failure *vis-à-vis* providing a standard of correctness and incorrectness for their members. Brandom provides no such thing. Future communities simply are not accessible in the way that they would have to be to provide such a standard. They are, in fact, in-principle *unavailable* to present communities for consultation.

Brandom's solution fails, then, for the same reason that Blackburn's proposal that time-slices of an individual can act as the standard for correctness and incorrectness of

that individual's judgments fails. Whatever is to act as this standard must be in place *compresently* with that which is to be corrected. This is the only way that a genuine correction can take place. The lesson we should take away from Brandom, then, is that if a coalition is to be the standard for the correctness and incorrectness of a single community, it must be—at least at some point—a synchronic coalition.

Putting this lesson together with our account of rational theory change, we arrive at the following picture. A single community may go along enforcing certain conceptual norms, postulating certain objects as explanations for the manifold of intuition and the necessities therein, etc. As it does so certain anomalies might crop up, certain inconsistencies between what the community predicts will happen with the manifold of experience and what actually happens. Such anomalies may be paid a great amount of attention by the community, or they may not. The community may reject its own conceptual scheme as inadequate, or it may not. As of yet, there is nothing about this behavior that makes it the behavior of a community that *represents* itself providing a standard of correctness and incorrectness for its members. This part of our story parallels the part of our story about the individual in which he goes about making various utterances with no real standard in place for their correctness or incorrectness. The community is in an exactly similar position for all that we have said so far.

What makes this behavior of the community's that of representing itself as providing a standard of correctness and incorrectness to its members is adding to this picture the fact that such a community would recognize the authority of another community with a conceptual scheme possessing the features presented in our first section as a *corrective of its own conceptual scheme*. That is, what makes our first

community a representing community is that it would recognize a conceptual scheme that explains not only the manifold of intuition, but also the reason why the previous conceptual scheme did as well as it did for so long, and why it ran into just the anomalies that it did as it went, as a better own than its own. Without such recognition the judgments of success and failure of the first community are arbitrary whims, held to no standard at all. With it, we have a coalition that is in a position to correct one another according to a very particular standard put in place by the very same rational need that provides the need for communities in general (the project of conceiving of oneself as a single, unified subject of experience persisting through time).

The agreement of this coalition that they have met *this* standard then becomes the standard to which each particular community is held. If a conceptual scheme is decided to be replaceable in the appropriate way by another conceptual scheme, then the first is represented as adequate to its task only *incorrectly*. The agreement of the coalition as to whether two conceptual schemes stand to each other in this relation acts as the standard for the correct and incorrect judgments of each community regarding their success or failure in providing a standard of correctness and incorrectness to their members. That is, what it takes for a community to *fail* to provide its members with a standard of correctness or incorrectness is just for their to be some other such standard, enforced by some other community, that a coalition of communities agrees *does a better job of this* that does that of the first community.

We have now arrived at the point where it seems natural to once again push a worry about consistency. The individual could not act as his own standard of correctness and incorrectness in part because he was prone to inconsistency. The community could

not act as the sole standard of correctness and incorrectness for the individual because *it* was prone to inconsistency. We have now suggested that a community of communities, or a coalition, can act as the final arbiter of correctness and incorrectness, so to speak, and it seems natural to wonder if it too is prone to inconsistency. Could it not happen that some such coalition could, at different times, amongst different factions, etc. declare one conceptual scheme both better and worse (more and less explanatory) than another? There certainly seems to be nothing that would rule this out a priori. Are we then, once again, in the position of finding yet another standard of correctness and incorrectness, this time for the coalition?

To begin to answer these questions it will be helpful to recall a moment from our last chapter, specifically, the moment at which we introduced to our lone individual another self-consciousness. We said there that the situation was one in which each saw the other as engaged in the same project as himself—experiencing and attempting to make coherent the world around him—and that as such each could recognize the other as on equal footing with himself. We considered there both the scenario in which the two self-consciousnesses agreed with one another and the scenario in which they disagreed. This was the point at which we moved to the considering the *eventual or ideal agreement* of the community formed by two such self-consciousnesses as the standard of correctness and incorrectness for each. That is, we had already confronted at that moment a community in contradiction, the community consisting of two *disagreeing* parties. Our solution there was to note that *if* these parties were to ever have agreement—which considerations stemming from the Transcendental Deduction showed they *must* each take as a goal—then that agreement could be taken as the standard of correctness and

incorrectness for each. That is, we suggested that coming to some agreement was a *regulative ideal* for each self-consciousness, and thus that they would each forge ahead until some such agreement was reached. What I want to propose now is that a similar solution is available here in the case of disagreement among the members of a coalition.

The cases of intra-coalition contradiction we are considering are each cases where some part of the coalition—be it a time slice of the coalition, or a particular community within the coalition, etc.—disagrees with some other part of it. This would seem to be an intolerable contradiction that threatens to undermine yet another proposed standard of correctness and incorrectness. What I want to push now is the suggestion that such a contradiction, at this level, is not vicious as it was earlier. That what we ought to take as the standard of correctness and incorrectness here is not the coalition as it is at *any* particular time, but only as it is when it reaches a point of agreement (about what conceptual schemes are better than what others). The suggestion at hand is further that such an agreement acts as a *regulative ideal* of the practices of such a coalition. That is, the communities that constitute the coalition must, on pain of giving up their *constitutive* goal of providing a standard of correctness and incorrectness for their members, strive towards agreement with their fellow communities. As Hegel puts it—because we have now moved again from Kant’s back to Hegel’s territory—

The reconciling *Yea*, in which the two ‘I’s let go their antithetical *existence*, is the *existence* of the ‘I’ which has expanded into a duality, and therein remains identical with itself, and, in its complete externalization and opposite, possesses the certainty of itself. (PS §671)

Hegel is here writing of what he would call the ‘I’ that is ‘We’, i.e., of two *communities*, not individuals. His point is just the same one that we saw him make five-hundred or so paragraphs earlier: that a standard for correctness and incorrectness is only available

when there exist both the possibility for disagreement and actual agreement. Each community achieves this standard by “the reconciling Yea,” by coming to agree with a community with which it previously, or merely possibly, disagreed (“its complete externalization and opposite”). Such agreement is a regulative ideal for each community, and so the constitutive goal of the coalition as a whole.

Since it is the *eventual* agreement of the coalition that is taken as a standard of correctness and incorrectness by the community, it is only this eventual standard that we need worry about running into contradiction. Since the sole purpose of the coalition is to judge which community has the most adequate conceptual scheme, and since it will use something like the guidelines outlined in the first section of this chapter to do so, it would be strange indeed for such a coalition to contradict itself. Contrast, for instance, this kind of contradiction with the ones we encountered earlier with the individual and with the community. The individual ran into contradiction when he attempted to apply concepts in a haphazard way to the objects in the world. The community ran into contradiction in its attempts to produce an adequate representation of the world. Each of these is fairly predictable. Neither the individual nor the community had a *procedure* on hand for avoiding contradiction. No univocal result ought to have been expected of them. The coalition, on the other hand, has exactly this.¹⁵³ Again, if something like the guidelines from our first section are up to the task, the coalition has a procedure for determining when one conceptual scheme is better than, and so ought to replace, another. Since it is *only* when some conceptual scheme meets these criteria that it will be adopted, the procedure as a whole can be seen as one of steady progress towards a final agreement.

¹⁵³Pace, Kuhn, (1977).

There is no *room* for contradiction here, and so the consistency of the agreement of the coalition is assured.

Such a position faces several immediate challenges that we will now try to meet. The first of these is the following. A coalition is meant to provide a standard of correctness and incorrectness for the various communities that constitute it. As we have pointed out on several occasions, this requires that such a standard be available to those communities. If, however, the *eventual* agreement of the coalition is what acts as this standard, it would seem that such a standard will be *unavailable* to all those communities that are no longer in place at the time of this agreement.

It is here that we must come to understand the role of the eventual agreement of the coalition as a *regulative ideal* in more detail. That this agreement functions as a regulative ideal for communities means that insofar as it is within their power, communities will work to bring it about, insofar as they do not do this, they can be properly criticized by other communities, and that this work is not complete until such an agreement *is* brought about. What this means, then, is that while the eventual agreement of the coalition can act as a standard of *correctness* for communities choosing conceptual schemes, until such an agreement is reached, *disagreement* among such communities can function as a standard of *incorrectness*, or *inadequacy*.¹⁵⁴ Communities, on this picture, seek to bring about agreement with other communities (through the proper procedures). Each uses as its standard of correctness whether or not it has yet brought about such an agreement. Thus, each community is correctable by all the others. No community's

¹⁵⁴ Again, such an account will have to be supplemented with an account of *justified* disagreement, *proper* procedure, etc. The guidelines for replacing conceptual schemes outlined in the first section of this chapter will, presumably, go some way towards filling out such an account.

conceptual scheme can be considered correct, until all communities come to agree that it is. It is a working goal of such communities to bring about such an agreement.

More concretely this situation is one in which, if the first section of this chapter was along the right lines, communities will work to bring about the eventual agreement *by* replacing their conceptual schemes in accordance with the guidelines there outlined. Those guidelines act as a means to the community's necessary end. There is a sense, then, in which the eventual agreement of the coalition acts as a standard for individual communities in each and every change of conceptual scheme. It is not the case that these communities *compare* their judgments with those of the coalition (as was roughly the case with the individual and the community). Rather, it is that each community takes *bringing about* an agreement of the coalition as its *reason* for changing its conceptual scheme according to the guidelines outlined above. The immediate corrective, then, is not the eventual agreement of the coalition, but rather the present *disagreement* between communities. Such a disagreement *must* be overcome, if the communities are to succeed in providing their members with demonstrably adequate conceptual schemes. It is by adopting a procedure like the one outlined in the first section of this chapter, that this can be achieved.

It is worthwhile to notice here the difference between the present proposal and the proposal of Brandom's that we earlier critiqued. On Brandom's proposal, the two-part standard to which each community is held is that of making their conceptual scheme consonant with both their predecessor communities and their successor communities. Our objection to such a picture was that there is no way for such a standard to put any *actual* constraint on the practices of the present community. There is no way, on that picture, for

the present community to know what the future community will make of its conceptual scheme, and so no way for the present community to *use* the future community *as* a standard. The present proposal avoids this worry. According to it, the present community need not know what the future community will make of its conceptual scheme at all to know whether it is correct or incorrect. What the present community needs to know, and can know, is whether there is agreement among all the communities that are compresent with it. If not, then there is still work to be done. The community cannot yet consider its conceptual scheme adequate.¹⁵⁵ Since it is compresent *disagreement*, once again, that is doing the work here, rather than future agreement, the present proposal can avoid the difficulties that Brandom encountered.

The next set of questions we must answer concern the analogy, or disanalogy, between the coalition and the individual, and the coalition and the community. That is, if our proposal resolves the contradictions of the coalition, we must explain why a similar solution is unavailable to the individual and the community. We can begin with the individual. Suppose, as we have previously, that we divide the individual by his temporal parts. The suggestion at hand would then be that while this group of the individual's temporal parts disagree amongst themselves, they might take it as a regulative ideal that some day they will agree, and that what they agree on in this ideal circumstance should act as the standard of correctness and incorrectness for each of them. At least two things should strike us as odd here. Firstly, since we are speaking of *temporal* parts, the notion of *eventual* agreement of such parts is a difficult one to parse. There can be no one *time* at which they all agree because *ex hypothesi* they all exist at different times. Secondly, there

¹⁵⁵If, on the other hand, some set of compresent communities *does* reach agreement, this may not yet be adequate for judging that the agreement of the coalition has been reached. There is always further evidence to gather, etc.

can also be no process by which all of these temporal slices come into agreement because no temporal slice has the *time*, so to speak, to change its mind. By the time it has judged one way, it ceases to exist, and certainly—as we noted earlier—cannot have its mind *changed* by a temporal slice that exists only when it does not. In sum, the individual’s temporal parts can come into a kind of disagreement with one another, but have no way of resolving this disagreement.

The trouble with the community, at least in the interesting cases at which we looked earlier, is that there is no *disagreement*. That is, the contradictions that we were considering earlier arose from the community’s agreeing that a certain judgment, which it endorses, implies that the manifold of intuitions should be a certain way. The contradiction arises when the manifold of intuitions is not that way. The community, we are supposing, recognizes its own failure. There need be no disagreement among its members at all. The “disagreement”, if there is any, is between the judgments of the community and the world. *This*, however, is not the kind of “disagreement” that is overcome. Rather, as we saw earlier, it is reason for abandoning the conceptual scheme that the community had been employing in favor of some other. The regulative ideal that the community would adopt in such a case, if we were to describe it as such, would be to use as a standard of correctness and incorrectness whatever conceptual scheme accorded best with the manifold of intuitions.¹⁵⁶ This is just to say, however, that the community adopts the practices that we have already described in this chapter: employing the

¹⁵⁶Of course, this supposes that there will be some agreement as to what “accords best” with the manifold of intuitions. Kuhn, for instance, would push that this is a mere pipe dream that we have no reason for thinking will ever happen. The line we have pushed above, however, is that the reason we have for thinking that a coalition *could* come to some such agreement is that all of its members take it as a regulative ideal that they *must* come to such an agreement. That is, the reason we should expect that the coalition will agree is that the coalition is set up so that it will not cease its activity until it has done so. It is entirely within our control, that is, to bring such an agreement about, and if the line we have traced from the Transcendental Deduction is right, it is *impossible* for us to give up this goal.

account of rational theory change that we have outlined in accordance with the judgments of the coalition. Thus, depending on how one wishes to speak, either the community *can* employ the solution we have outlined for the coalition *by* forming such a coalition, or it cannot employ this solution because this solution is a way of resolving disagreements of which the community has none.

BIBLIOGRAPHY

- Bennett, J. *Locke, Berkeley, Hume: Central Themes*. Oxford: Clarendon Press.
- Berkeley, G. (1998) *Three Dialogues Between Hylas and Philonous*, ed. Jonathan Dancy. (Oxford: Oxford University Press).
- Blackburn, S. (1984) "The Individual Strikes Back", *Synthese*, 58, pp. 281-302.
- Brandom, R. (1994), *Making It Explicit*. Cambridge, MA: Harvard university Press.
- (2002), *Tales of the Mighty Dead*. Cambridge, MA: Harvard university Press.
- Brandt, R. (1995), *The Table of Judgments: Critique of Pure Reason A67-76; B92-101*. Trans. Watkins, E. Atascadero, CA: Ridgeview Publishing Company.
- Brittan, G. (1978), *Kant's Theory of Science*. Princeton: Princeton University Press.
- Carruthers, P. (1983), 'On Concept and Object', *Theoria* 49: 49-86.
- Descartes, R. (1984), *The Philosophical Writings of Descartes*. Volume 2. Trans. Cottingham, J., Stoothoff, R., & Murdoch. D. Cambridge: Cambridge University Press.
- Dicker, G. (1998) *Hume's Epistemology and Metaphysics*. London: Routledge.
- Dretske, F. (1986) "Misrepresentation", in *Belief: Form, Content, and Function*. Ed. Bogdan, R. J. (Ed.) (New York: Oxford University Press).
- Evans, G. (1982) *The Varieties of Reference* (Oxford: Clarendon Press).
- Everson, S. 'The Difference Between Feeling and Thinking', *Mind* 37: 401-13.
- Fodor, J. (1990) *The Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press).
- (2003), *Hume Variations*. Oxford: Clarendon Press.
- Forster, M. (1998) *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press).
- Garrett, D. (2002), *Cognition and Commitment in Hume's Philosophy*. New York: Oxford University Press.

- Gibson, M. (2004), *From Naming to Saying: The Unity of the Proposition*. Malden, MA: Blackwell Publishing.
- Hanna, R. (2001), *Kant and the Foundations of Analytic Philosophy*. Oxford: Oxford University Press.
- Hegel, G. (1976) *Science of Logic*. Trans. A.V. Miller. New York: Humanities Press.
- (1977) *Phenomenology of Spirit*, trans. Miller, A.V. (Oxford: Oxford University Press).
- Henrich, D. (1989), ‘Kant’s Notion of a Deduction and the Methodological Background of the First *Critique*’, in E. Förster (ed.) *Kant’s Transcendental Deductions*. Stanford: Stanford University Press.
- Hume, D. (1978), *A Treatise of Human Nature*. L. A. Selby-Bigge (ed.) New York: Oxford University Press.
- (2002), *A Treatise of Human Nature*. D. Norton and M. Norton (eds) New York: Oxford University Press.
- Everson, S. (1988) “The Difference Between Feeling and Thinking.” *Mind*, Vol. 37, No. 387: 401-13
- Fodor, J. (1990), *The Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Fodor J. and Lepore E. (1992), *Holism: A Shopper’s Guide*. Oxford: Blackwell.
- Kant, I. (1998), *Critique of Pure Reason*. Trans. P. Guyer & A. Wood. Cambridge: Cambridge University Press.
- Kemp, C. (200) “Two Meanings of the Term “Idea”: Acts and Contents in Hume’s “Treatise””, *Journal of the History of Ideas* 61: 675-90.
- Kuhn, T. (1977) “Objectivity, Value Judgment and Theory Choice”, in *The Essential Tension* (Chicago: University of Chicago Press).
- Linsky, L. (1992), ‘The Problem of the Unity of the Proposition’, *Journal of the History of Philosophy* 30: 243-73.
- Longuenesse, B. (1998), *Kant and the Capacity to Judge*. Trans. Charles T. Wolfe. Princeton: Princeton University Press.
- Millikan, R.G. (1984) *Language, Thought and Other Biological Categories*. (Cambridge, MA: MIT Press).
- Noonan, H. (1999) *Hume on Knowledge*. London: Routledge.

- Ott, W. (2004), *Locke's Philosophy of Language*. Cambridge: Cambridge University Press.
- Pippin, R. (1982), *Kant's Theory of Form*. New Haven, CT: Yale University Press.
- (1989) *Hegel's Idealism* (Cambridge: Cambridge University Press).
- Pinkard, T. (1996) *Hegel's Phenomenology* (Cambridge: Cambridge University Press).
- Rosenberg, J. (1968), 'Wittgenstein's Theory of Language as Picture', *American Philosophical Quarterly*, 5: 18-30.
- (1980), *One World and Our Knowledge of It*. Dordrecht, Holland: D. Reidel Publishing Co.
- (1986) *The Thinking Self*. Philadelphia: Temple University Press.
- (2005), *Accessing Kant: A Relaxed Introduction to the Critique of Pure Reason*. Oxford: Oxford University Press.
- (2007), *Wilfrid Sellars: Fusing the Images*. Oxford: Oxford University Press.
- Seibt, J. (1990), *Properties as Processes: A Synoptic Study of Wilfrid Sellars' Nominalism*. Atascadero, CA: Ridgeview Publishing Company.
- Sellars, W. (1953), "Inference and Meaning" *Mind*, 62, 247: 313-338.
- (1963a), 'The Language of Theories', in *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company, (1963), 106-28.
- (1963b), 'Naming and Saying', in *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company, (1963), 225-46.
- (1963c), 'Some Reflections on Language Games', in *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company, (1963), 321-358.
- (1963d), 'Empiricism and Abstract Entities', in *The Philosophy of Rudolph Carnap*. Ed. P.A. Schilpp. ; LaSalle, IL: Open Court.
- (1964), 'Induction as Vindication', *Philosophy of Science*, 31: 197-231.
- (1967), *Science and Metaphysics*. Atascadero, CA: Ridgeview Publishing Company.
- (1979), *Naturalism and Ontology*. Atascadero, CA: Ridgeview Publishing Company.

- (1988) “On Accepting First Principles.” *Philosophical Perspectives*, 2 (1988): 301-14.
- (2002a), ‘Towards a Theory of the Categories’, in *Kant’s Transcendental Metaphysics*. Atascadero, CA: Ridgeview Publishing Company.
- (2002b), *Kant and Pre-Kantian Themes: Lectures by Wilfrid Sellars*. P.V. Amaral (ed.) Atascadero, CA: Ridgeview Publishing Company.
- Strawson, P. F. (1959), *Individuals*. London: Methuen.
- Stroud, B. *Hume*. (1977) London: Routledge and Kegan Paul.
- Wiggins, D. (1984) ‘The Sense and Reference of Predicates: A Running Repair to Frege’s Doctrine and a Plea for the Copula’, *Philosophical Quarterly*, 34: 311-28.
- Wittgenstein, L. (1961), *Tractatus Logico-Philosophicus*. Trans. Pears, D. F. & McGuinness, B. F. London: Routledge & Kegan Paul.
- (1953) *Philosophical Investigations*, trans. Anscombe, G. E. M. (Oxford: Basil Blackwell).
- Wright, C. (1981) “Rule-Following, Objectivity and the Theory of Meaning”, in Holtzman, S. & Leich, C. (Eds.), *Wittgenstein: To Follow a Rule*, pp. 99-117 (London: Routledge & Kegan Paul).