Jackson—Week 5

The Secrets of A-intensions Revealed
Or
How to Hold Your Naturalism Hostage to an Implausible Empiricism

1) <u>Constraints on A-Intensions</u>, given the theoretical role Jackson wants them to play. There are two principle constraints (corresponding to the two theoretical roles we came in the 70s to realize were played by Fregean *senses*). I'll include also a third, which Jackson thinks is a consequence of the second:

a) *Semantic Determination*: The A-intension of a term *t* determines, at any world *w* ("treated as though it were the actual world") both an extension (treating *w* as though it were the actual world ensures that the C-extension at *w* and the A-extension at *w* will coincide, even though the C-intension determined at *w* by the A-intension in question may—and for two-dimensional terms in general will—determine C-extensions at other worlds *w'* that differ from the A-extensions at *w'*) and a C-intension for *t*. The idea is that the A-intension of an expression *t* is whatever *fixes the reference* of *t*. As we take that common reference-fixer (a "descriptive content") from world to world, it fixes different referents, depending on the facts in that world. On an XYZ world, what we would have been talking about with our term 'water'—the colorless, odorless liquid we drink and bathe in and find in lakes and streams—would have been XYZ rather than H_2O.

b) *Dual Independence*: The A-intension of a term does not depend on either:
    i) Which world is actual, or (therefore)
    ii) The actual C-intension of the term *t or of any other terms*. (For if it did, the A-intension would depend on which world is actual, for one needs to settle that to settle the actual C-intension.)

The idea here is that one does not need to settle what world one is in to settle what fixes the reference of a term. One needs to settle what world one is in only to settle what referent is actually fixed by that reference-fixer.

c) *Epistemic Transparency*: One can *know* what the A-intension of *t* is without knowing what world one is in, that is, without any knowledge of contingent facts. Hence one can grasp A-intensions *a priori*. Jackson argues that this is a consequence of *Dual Independence*: since one does not need to *settle* (determine) what world one is in to *settle* what fixes the reference of a term, but needs to *settle* what world one is in only to *settle* what referent is actually fixed by that reference-fixer, one does not need to *know* what world one is in to *know* what fixes the reference of a term, but needs to *know* what world one is in only to *know* what referent is actually fixed by that reference-fixer.

Although I will not contest Jackson's claim that *Epistemic Transparency* follows from *Dual Independence*—a claim that turns on or just expresses his definition of "*a priori*" (which is not the only definition one might offer, but is certainly an acceptable thing to mean by "*a priori*", construing it in terms of possible worlds)—I think it is worth keeping separate track of this constraint. By Jackson's own lights, it is at least a *necessary* condition of *Dual Independence*, but it might be satisfied though that constraint was not.

It will turn out that it is harder than one might think to describe A-intensions so that they satisfy all three of these constraints—or so I shall argue.

2) <u>On the internal structure of A-intensions</u>: It turns out that quite a lot can be said about the internal structure of A-intensions.

a) As Jackson conceives them, the A-intensions of expressions correspond to *theories*: folk theories, which codify *intuitions about possibilities* of those who use those expressions. (He says that such intuitions are "…[the engine that drives…] [ref.] [get proper quote from around 55]".) His thought is that one finds out what the concept corresponding to '*t'* is by eliciting intuitions about possibilities, and then codifying or elaborating those into a mini-theory about 't': for instance, that water is the colorless, odorless liquid we drink, and bathe in, and find in lakes and streams. *That* is then what is carried over from world to world as the A-intension. And *that* is what can be grasped *a priori*, without one's having to know what it is in the actual world that that theory is true of.

b) How does the folk-theory associated with the term *t* determine, at some world *w*, what *t* refers to at that world (in the sense of what it *would* refer to *if w* were the *actual* world)? Jackson says it is whatever "plays the watery role" in *w*. [ref.] He gives us some help with this by invoking David Lewis's (and others') discussion. Although he does not use the term, the technique of theirs that he is appealing to is *Ramsification* of a theory.

c) Explain Ramsification: Suppose θ is a theory in some first-order language L, in the sense of a set of sentences of that language. We may designate by T the set of logically primitive non-logical terms (individual constants and predicates, but we will focus on the predicates) that occur in the sentences of θ. The *Ramsey sentence* that expresses the role played by any particular predicate P_i in θ is then what one gets by:

i) Forming the conjunction $C(\theta)$ of all the sentences in θ;

ii) Replacing all occurrences of P_i by a predicate (second-order) variable X_i to produce the second-order one-place predicate $C(\theta)/P_i$;

iii) Prefacing $C(\theta)/P_i$ by an existential quantifier, to yield $\exists X_i[C(\theta)/P_i]$.

The idea is that the second-order, Ramsified predicate $C(\theta)/P_i$ expresses the *role* played by P_i in θ. The corresponding Ramsey sentence just, says that *something* plays that role. Any world in which that Ramsey sentence is *true* is a world in which something plays the P_i (e.g. 'water') role. So we go from the folk mini-theory (already in effect in conjunctive form):

θ: "Water is a colorless, odorless liquid we drink, and bathe in, and find in lakes and streams." to the Ramsey sentence $\exists X_i[C(\theta)/P_i]$: "There is something that is a colorless, odorless liquid we drink and bathe in and find in lakes and streams."

d) This may seem not to be much of an advance. But notice that we can *repeat* the process with other terms in the original theory. So we can get the Ramsey sentence $\exists X_i,X_j[C(\theta)/P_i,P_j]$: "There are things X_i and X_j such that X_i is a colorless, odorless X_j we drink, and bathe in, and find in lakes and streams. This is a sentence that can be true in a world unlike our own not only in that it has XYZ in it instead of H_2O, but also in that instead of our familiar substance-phases solid, liquid, gas, and plasma, it has somewhat different ones—say, pliquids [the 'p' is

silent, as in 'swimming'], which are like our liquids, except that if shaken vigorously enough, they freeze (turning into psolids).

e) Indeed, we can carry the Ramsification process all the way to the limit, in which we have replaced *all* the non-logical terms in the theory θ by existentially quantified second-order predicate-variables. What is left is a *purely structural specification* of the theory—or, if you like, a specification of the purely logical structure of the theory, in which we have abstracted from all the *content* of the non-logical predicates *save what content they get from their functional relations to each other according to the theory*. Lewis wanted to do this for an (imaginary, of course) *psychological* theory that specifies what someone is thinking, believing, desiring, and so on. Then his way of expressing a functionalist theory of the psychological is to say that anything of which the purely structural Ramsey sentence corresponding to that theory is true is something that is having those same thoughts, beliefs, and desires, no matter what the material or other nature of the 'realizers' of those functional roles is—that is, no matter what it is in a given world that makes the pure existentially quantified second-order Ramsey sentence true.

f) [Aside for those who like to think about this sort of thing: Beth's theorem specifies a particular sense in which one set of non-logical predicates in a first-order theory can be said to *implicitly define* the rest (when all models of the theory that agree on the interpretation of the first or base set also agree on the interpretation of the rest). And of course the theorem then tells us that whenever that is true (a model-theoretic thing one could mean by 'supervenience'), the terms that are implicitly definable in this sense can also be *explicitly* defined. What happens if you Ramsify just a set of predicates that are in this sense implicitly defined by the rest? What relation do such Ramsey sentences bear to the original theory?]

g) The point of all of this in Jackson's case is that we can think of what we carry over from world to world as reference-fixing A-intensions as Ramsifications of the folk theories that codify the conceptions elicitable from the intuitions about possibilities evinced by the folk.

3) We have seen that the Ramsey sentences corresponding to theories can have more or fewer of their non-logical predicates Ramsified, by being replaced by existentially quantified second-order predicates. Some Ramsey sentences still contain a substantial residue of non-logical expressions, while the pure Ramsey sentences have eliminated all such non-logically contentful expressions. Where, on this spectrum, should we think of the Ramsey sentences corresponding to A-intensions of various expressions: the ones that codify the 'watery role', for instance?

a) One possibility, at least in the abstract, is that A-intensions correspond to *pure* Ramsifications of the folk theories elicitable from the intuitions about possibility of the folk. This alternative would have the advantage that it is very plausible that the conditions of *Dual Independence* and (so) *Epistemic Transparency* are satisfied by such purely structural specifications of functional roles. For there is no non-logical content appealed to in pure Ramsey sentences. *All* that is left is the relations between the various non-logical predicates that can be specified in logical terms. That purely logical structure owes nothing to any particular world

or set of contingent facts. It is available to be considered no matter what world we are in, and is in that modal sense intelligible purely *a priori*.

b) The questions we must ask about this alternative, then, is whether A-intensions corresponding to pure Ramsifications of folk theories can satisfy the *Semantic Determination* constraint. Do they suffice, when put together with an arbitrary possible world *w* ("thought of as the actual world") to determine extensions for the Ramsified terms, and C-intensions that will let us follow them from world to world? [See the note below on rigidification of fixed reference and the Kaplan 'dthat' operator.]

c) I think it can be shown that the *Semantic Determination* constraint can*not* be met by pure Ramsifications of folk theories.

i)   The problem is not that the pure Ramsey sentences corresponding to such theories are not *true* at arbitrary worlds. They are.

ii) The problem is rather that they are *promiscuously* satisfied. Pure Ramsey sentences have abstracted away all non-logical terms whose intended interpretations might constrain our choice of models for them. So *all* we need for the Ramsey sentence to be *true* at a world *w* is for there to be *some* way, *any* way to find in that world a *domain* of objects and a *set of relations* (one for each of the non-logical predicates P_i that we have replaced by existentially quantified second-order predicate variables) such that the abstract relational structure consisting of that domain and those relations is a *model*, in the Tarskian model-theoretic sense, that satisfies the pure Ramsey sentences.

iii) But it is *easy*—*too* easy—to pick out such a domain and define on it a set of relations that will satisfy an arbitrary given pure Ramsey sentence. For, given any region of space, say that occupied by a chunk of iron, I can subdivide it up into spatial regions (indeed, into an infinite, even uncountably infinite number, if need be) and use those for the domain. And then we can simply gerrymander sets of those regions into sets of n-tuples, reading off of the theory how the relations should be defined on our domain of spatial regions. Put another way, if there is *any* model of the pure Ramsey sentence, in the model-theoretic sense—which is just to say, if the sentence is *consistent*—we can map the domain of *that* model one-to-one onto the spatial sub-regions we have carved out of our initial space, in any arbitrary fashion, and then use that same mapping to translate the sets of sets of domain elements that are the relations in the model into sets of sets of spatial sub-regions in the isomorphic model we are constructing in possible world *w*. The lesson is that structure comes very cheap if there are no constraints on the interpretation function that maps elements of the pure Ramsey sentence onto items in the possible world.

d) This point, by the way, has long been appreciated (though it is not always kept firmly in mind) by functionalists in the philosophy of mind who are looking for 'realizations' of functional architectures. If a functional system is described abstractly enough, *any* physical system can be interpreted as realizing it. The lesson they have drawn, quite appropriately, is that you have to leave *some* non-logically contentful predicates in the Ramsey sentence specifying the functional architecture, if there is to be any line drawn between what does and does not realize that architecture, exhibit that functional organization. So, for instance, it is

a good thing if *time* references are held fixed as indicating *times*—rather than, say, distances or temperatures. Another popular candidate for even minimally substantive functional specifications is *causation*. For if we Ramsify that notion, 'causes' may be mapped onto the relation of mere spatial proximity of regions. This is the lesson we should draw in the case of A-intensions too.

e) In the first segment of the course, I underlined the point that the possible-worlds framework and the model-theoretic frameworks are *different*, and not straightforwardly translatable one into the other. I suggested that for these reason one should be very cautious about transferring results from one framework to the other—for instance, drawing conclusions about supervenience, thought of in possible-worlds terms, from Beth's model-theoretic definability theorem. We saw Stalnaker (in his Appendix) offering a particularly egregious example of an incautious slide between these frameworks. And I mentioned Etchemendy's treatment of the concept of logical consequence as an example of a substantial philosophical argument that turns on being very careful indeed about the relations between arguments conducted in these two frameworks. The argument I just presented makes some more positive claims about the relations between the frameworks, highlighting one of the principal sources of potential strain between the two: the fact that models do, but possible worlds do not, as such come with *domains*.

f) In the present context, the result of the promiscuous satisfaction of pure Ramsey sentences in arbitrary possible worlds is that no *unique* extensions, and hence no unique C-intensions, are determined by the pure Ramsifications of any folk theories. Pick any object, property, or relation in any world, any pure Ramsey sentence, and any Ramsified term in it (of the proper adicity), and it is easy to whump up an interpretation according to which that bit of the possible world is the extension of the term or predicate. The pure Ramsey sentence gives us so little to work with that if it is all we have to rely on, the reference of any term could be fixed to anything. Pure Ramsey sentences are semantically completely *non*-determining.

4) So we can see that we cannot Ramsify wildly if we want to satisfy the *Semantic Determination* constraint. *Some* non-logical expressions will have to remain unRamsified, so that their having to take their intended interpretations will constrain what counts as making the Ramsey sentence codifying the folk theory true, in turn constraining what extensions can be assigned to the *other* terms, which *are* Ramsified.

a) Of course, if we don't Ramsify *any* of the terms, but leave *all* the non-logical terms as having to take their intended interpretations (a catch-phrase of the *application* of model theory, though not of the mathematical theory itself), then all the terms will take the references they have in the actual world, no 'watery roles' will have been specified that could be played by XYZ rather than H_2O, and hence no A-intension satisfying the constraint of *Dual Independence* will have been specified.

b) The question then becomes how to partition the vocabulary in the folk theory into two mutually exclusive and jointly exhaustive classes:

i) the expressions we do Ramsify, forming roles that can then be played by, and so fix the reference of the corresponding expressions to, other things in other worlds, on the one hand, and

ii) the expressions that will form the non-logical core or backbone of the Ramsified folk theory, the bits that get a non-arbitrary grip on bits of other possible worlds, making it possible for the whole thing to determine references for the remaining expressions in arbitrary possible worlds.

c) The predicates forming the unRamsified core of the Ramsey sentences representing folk theories, defining the *P-role* for various values of P, can themselves, of course, be any A-intensions. But if these are themselves Ramsey sentences, we face the same choice as before: either they are pure or not. If they are pure, they can't do the job of contributing to the reference-fixing job in other worlds that *Semantic Determination* requires. If they are not, we are just put off to the non-logical predicates that appear as *their* core. (I don't *think* it makes any difference whether we think of *substituting* other impure Ramsey predicates—predicates formed by partially Ramsifying other theories—for the core non-logical predicates in a partially Ramsified theory, say, plugging in a Ramsified 'lake'-role for the predicate 'lake' in the partially Ramsified specification of the 'water'-role, as opposed to simply forming a more comprehensive theory, all at one level, and then partially Ramsifying it. But this is actually a substantive formal issue: there is a fact about whether any way of doing one has an equivalent of the other form, and I have not thought it through sufficiently to be confident what that fact is.) So in any case we are going to be driven to the conclusion: Not *all* A-intensions can have the form of Ramsified folk theories, because

i) Pure or completely Ramsified theories are too thin to fix references and hence satisfy the *Semantic Determination* constraint on A-intensions, and

ii) Incompletely Ramsified theories, which *do* have non-logical content in addition to their structure that at least potentially might suffice to fix reference do so because they incorporate *un*Ramsified non-logical predicates that cannot themselves, on pain of an infinite regress, be understood as having the structure of (predicates derived from) Ramsey sentences.

d) Those *basic* A-intensions, the ones that give content to the Ramsey specifications of roles of *other* terms, must themselves meet the three conditions of *Semantic Determination*, *Dual Independence*, and *Epistemic Transparency*. That is, for them to play the role they must in more complex A-intensions formed from them by Ramsification of folk theories in which they occur,

i) they must be followable to other worlds, in each of which, given that world ("taken as the actual world") they determine a reference;

ii) *what* we follow from world to world must not depend on any contingent fact about the actual world, or (therefore) on the C-intensions of any other expressions; and

iii) we must be able to grasp the content that fixes the reference in every world in (i) *a priori*, without knowing what world is actual (what world we are in when we consider other possibilities), that is without knowing anything at all about what (else) is *true*. So we can use *no auxiliary hypotheses* or *collateral premises* in making sense of these predicates.

e) These *basic* A-intensions *can* depend on the *A*-intensions of other expressions. Ultimately, this means that we *can* appeal to inferential connections *among the basic* A-intensions. (Relations to Ramsey-structured A-intensions must be resolvable into these.) That is, we need not be *atomistic* about the contents of these basic A-intensions. They may form a holistic set, with many only intelligible in terms of their relations to others. But the *whole set* of *basic* A-intensions must be intelligible, and intelligible as fixing reference (given a world), completely apart from their relations to *any* C-intensions, and apart from any assumptions at all about what is (contingently) true.

f) This set of basic A-intensions forms a *semantic foundation* for all semantic content. For it is in terms of these contents that the contents of all A-intensions that have the form of partially Ramsified folk theories are to be understood. And those A-intensions, together with worlds, in turn fix references and so (again, see the remarks about rigidification) C-intensions. So they ultimately determine all the A-intensions and (relative to worlds) all the C-intensions. And on the two-dimensional theory, that is all the content there is. Notice that this *semantic* foundationalism is not itself a form of *epistemological* foundationalism. It may lend aid and comfort to such a claim, given the *a priori* nature of our acquaintance with the basic A-intensions, but the fact that they settle the C-intensions in terms of which facts are stated only in the context of how the worlds are means that we cannot even know *a priori* what we are *referring* to, never mind what is true. So there is no immediate epistemological consequence for our *knowledge* of contingent matters of fact.

g) What could these basic A-intensions be?

5) One way of thinking about the question "What could the *basic* A-intensions be?" is to think about what vocabulary we the theorists are allowed to use in specifying them or otherwise picking them out.

a) We can, I think, use *indexical characters* in Kaplan's sense to do so. The character/content distinction for indexicals is one of the principal models from which the A-intension/C-intension distinction is elaborated. (What I am shaping up to argue for is that that model does not combine smoothly—as it must, for two-dimensionalism to do the work Jackson wants it to do—with the model of what fixes reference and the intensions that are thereby also fixed.) 'Here' has the character (function from indices to contents) *place of utterance*, and that, together with a world and an utterance in that world, settles the place referred to by that utterance in that world—a place that can then be rigidly traced through other worlds. The content fixed is a function from worlds to places, that assigns each world the *same* place: the place of utterance in the reference-fixing world). Similarly for 'now', and 'I'. More complex indexicals (such as "my mother's birthday") will be alright if and only if the other expressions they contain, the functions applied to simple indexicals, also express legitimate basic A-intensions). Lewis treats 'actual' as an indexical, which picks out the *world* of utterance, just as 'now' does the *time* of utterance and 'I' the *utterer*. This idea is a significant part of the motivation of two-dimensionalism: just follow uses of 'actual' from world to world. This is the cash behind the idea of "treating world *w* as the actual world."

b) It is clear that we cannot use *proper names* to specify basic A-intensions. For what the proper names we use refer to depends on what world we are actually in, what we are actually hooked up to by the chains of historical usage that link us with what our names are names of. So there can be no element of basic A-intensions (and hence, given how Ramsification works, no element of Ramsified versions of theories that are based on those basic A-intensions) that works to articulate its content in the way proper names do.

c) For similar reasons, we cannot use *demonstratives* to pick them out, or, therefore, think of them as incorporating demonstrative elements. It is tempting (and many philosophers succumb to the temptation) to assimilate demonstratives to proper simple indexicals, by thinking of an *demonstration* accompanying an utterance as being an index associated with it, in the same sense in which the time, place, speaker, or world (for 'actual') of utterance is an index associated with it. Being an indexical, after all, is just being such that the reference of an utterance is determined by a uniform function from some index associated with the utterance (not necessary for 'Aristotle' or 'red'). But the assimilation of deixis to indexicality is a serious mistake. For some feature of an utterance to be available to serve as an *index* in a semantics of this sort, it must be possible to determine the index associated with an utterance *independently* and *in advance of* determining what it refers to. And that condition is typically not satisfied in the case of demonstratives. [Tell this story.]

d) Anaphora

e) (thence to secondary quality, thence to phenomenal properties).

6) Descriptive Expressions:
   a) What about 'elm'? That is a descriptive predicate/sortal. But we need an A-intension that will:
      i) Distinguish between the extension of <u>elm</u> and <u>beech</u> in *this* world, as well as
      ii) at least *almost* all other worlds. (Is it possible that elms and beeches should have been the same? Certainly the *words* could have been used for the same tree-kind. But in *that* sense, they could also be *horses*.)
   b) Can I do that without appeal to *experts*? Surely *I* don't know enough to tell them apart in other possible worlds. (Well, I might, actually, but Putnam and Fodor are sure *they* don't, and I bet they are right. And there are a *lot* of other terms where I am in their boat.) Here, it seems, *folk* theories at the first level won't do, while folk theories that involve acknowledging that others know a lot more about elms than we do might.
   c) So suppose that the folk-theoretic A-intension of 'elm' includes something like "Elms are what the elm-experts say are elms." It seems that this could be understood in two ways [are there others?]:
      i) It is an instance of "Ks are what those who know a lot about K's say they are." The trouble with this horn of the dilemma is that we can say *exactly* the *same* thing about beeches, and, indeed, about molybdenum, and a *lot* of other semi-technical natural kinds. Satisfying *this* formula will not give a *unique* referent, even fixing worlds, to 'elm', which the A-intension must do, by *Semantic Determination*.

ii) It requires a substantive notion of <u>expert</u>. That could be a *generic* notion, of which <u>elm expert</u> is a species. But if we appeal to such a notion, we need an A-intension for <u>expert</u> that satisfies all three of our constraints, and which is either itself basic, or is resolvable into basic A-intensions via Ramsification of some *folk theory* of experts. Is <u>expert</u> going to be easier to do this with than <u>elm</u>? (I don't think it is just silly to ask whether we need to appeal, at least in particular cases, to experts on who is an expert. That's where I start when trying to choose a doctor/lawyer/accountant for some thorny problem.) Remember, the A-intension of <u>expert</u> is going to have to be available *a priori*, independent of any contingent matters of fact (e.g. about how reliable any individual as a matter of fact is, though it can appeal to reliability in general, *if* we can resolve *that* notion into basic A-intensions), and determine for any possible world who is an expert there on what.

d) This line of thought makes it likely that semi-technical descriptive natural-kind terms—those in which the folk are *not* experts—are not good candidates for *basic* A-intensions.

7) What are good candidates for descriptive terms that could be used to specify basic A-intensions, then?

a) We have seen that for *some*, but only some concepts, their A-intensions can be specified in terms of (Ramsifications of) folk theories. Thus Jackson says: "…the A-intension of 'temperature in gases' is that which plays the temperature role in gases in that world." [58]. But there must also be some non-logical expressions in terms of which such theories are articulated. On pain of a regress, *they* can't be understood as having the same sort of A-intensions, ones, namely, articulated by theories. We need some *non-theoretical* A-intensions.

b) 'Elm' and 'beech', and even 'molybdenum', are perfectly good *observational* terms for the suitably trained (the at least semi-expert), though not for the folk (us) in general. But we can look at expressions (of concepts) that are much more widely observable. (Note that Quine defines observationality as coming in degrees, depending on how widespread, over the population and differing collateral circumstances, the capacity to agree responsively on a verdict "given concurrent stimulation" is.) Jackson explicitly considers concepts that we do *not* have theories about, concepts that we can just reliably apply *responsively*, without being able to say how we distinguish Ks from non-Ks. He considers a distinction, appealed to by Dummett and Evans, between "descriptive identification" and "mere recognitional capacities". For it seems that descriptions must appeal to some concepts, the ones in terms of which something is described, and they can't themselves, on pain of a regress, all be descriptive. But Jackson wants to use 'descriptive' more generically, to include even the concepts whose use is based on intuitively non-descriptive "mere recognitional capacities." He says [66]: "being disposed to be recognized by a subject as such-and-such is a descriptive property of something that can and does serve to pick it out from other things."

c) The suggestion is that the description fixing the reference of, say, 'red' might be something like: "being such as to dispose suitably situated subjects reliably to respond differentially by taking it to be red." If, in any world *w*, there is

something that satisfies this description—that makes that Ramsey sentence true—then that is what 'red' gets its reference fixed to in that world.

d) I see two difficulties with this sort of specification of basic A-intensions:

i) One concerns the *response*. What is "taking it to be red"? If the A-intension of *that* expression were something like "applying the word 'red' to it"—so that *all* we took from world to world in following the A-intension around were the *sign design* or *noise*, then there are not only worlds in which 'red' means <u>green</u>, but worlds in which it means <u>heavy</u>, <u>slow</u>, <u>witty</u>, <u>serendipitous</u>, and so on. For the *noise could* have been used to mean anything at all. The basic A-intensions cannot just be *words*, on pain of utterly failing to fix 'red' in each world to things that are *red* in that world. The *response* that defines this *response-dependent* concept must itself have a *more basic* A-intension: one that can be followed from world to world *without* having to appeal to the notion of *what* it is a response *to*. This will have to be something like "seeming or looking red", where that is thought of as expressing a *phenomenal* property. And it will be a criterion of adequacy of the understanding of such properties that they are intelligible in complete independence of what actually *is* red at any world, in order to satisfy the *Dual Independence* constraint. Those familiar with Sellars's "Empiricism and the Philosophy of Mind" will see that "looks-red" cannot be given the analysis he offers there, if it is to satisfy that constraint. The Myth of the Given looms. I'll return to this thought below.

ii) The other difficulty concerns the rest of the terms articulating the response-dependent description used to pick out what is evidently being interpreted as the *secondary quality* <u>red</u>. We need something playing the role of the qualification "suitably situated"—we might talk about "standard conditions for observation", or something like that—because red things do not reliably dispose even competent observers to call them red if the observers are hundreds of miles away, blindfolded, in the dark, facing the wrong way, and so on. So without some such element in the description specifying the A-intension, the reference of 'red' will either not get fixed at all, or will be fixed wildly wrongly. But does the notion of <u>standard conditions for visual observation of colors</u> (which we can surely envisage changing its extension from possible world to possible world) itself plausibly have a *basic* A-intension? It is surely not a *phenomenal* property. If not—and the fact that what are standard conditions for visual identification of colors might be different in other possible worlds argues that it cannot be basic—what *are* the *basic* A-intensions in terms of which *this* description is articulated? The prospects for spelling out the notion of <u>standard conditions for visual observation of colors</u> in *purely phenomenal* terms are not promising. As Sellars argues in his essay "Phenomenalism", to spell out objective properties in terms of conditionals about phenomenal experiences, one needs to use conditionals whose *antecedents* themselves are couched in objective terms. It just is not the case that if I *seem* to turn the object around I will *seem* to see such-and-such, but that if I *actually* do so I will seem to see its hidden side.

iii) And things only get worse when we consider the rest of the response-dependent description. What is the A-intension of <u>reliable differential responsive disposition</u>? <u>Disposition</u> is a *modal* concept, a matter of what *would* happen if…. Dispositions, possibilities, and necessities are notoriously (especially among empiricists) not paradigmatically *directly observable* items. Are they to be taken to have *basic*, non-descriptive A-intensions? Or must we look to *folk theories* of reliability and counterfactual dependence? If so, what are the *basic* A-intensions in terms of which those theories are expressed? Empiricists ever since Hume have despaired of saying what 'necessary' and 'possible' mean in *phenomenal*, *experiential*, or even purely *observational* terms. What fixes the reference of these *modal* terms. (One of the central points of Sellars's "Counterfactuals, Dispositions, and the Causal Modalities", which we will read in the third quarter of the course is that modal talk is *not* just more descriptive talk, but articulates essential features of the *conceptual framework* that makes description is possible in the first place. Description and explanation are much more intimately bound up with one another than empiricism admits. But we'll pursue that point later.)

e)

8) The A-intension of <u>cause</u>:

a) Worrying about the A-intension of *modal* concepts such as <u>disposition</u> brings us to a possibly more fundamental difficulty. I hinted above that I think that the notion of A-intension results from the interesting, but ultimately unworkable idea of extending the model of the Kaplanian *character* associated with genuinely indexical expression types to Kripkean reference-fixing generally. The A-intensions of the two-dimensionalist are thought of as the *characters* associated with expressions generally. They, together with a world ("thought of as the actual world"), fix the reference and hence (see the remarks on rigidification below) the content associated that expression in that world. Indexical characters can do that with no help from anything else, as A-intensions are also required to do. Descriptions (such as "red ball") also are intensions that can be thought of as constant across worlds, picking out in each whatever satisfies them. (We've seen that this depends on their component concepts having constant intensions.) The two-dimensional framework would work best if all expressions behaved descriptively (including, paradoxically, the elements that articulate the descriptions).

b) Kripke has convinced us all that "descriptive theories of reference" are *not* correct for proper names, demonstratives, and at least some natural kind terms. (In fact, *no-one* ever thought that descriptive theories of reference applied to *all* expressions. Even raving descriptivists realized that the basic elements out of which the descriptions were formed had to have *their* reference fixed in some other way, on pain of an infinite regress. If you had asked them whether that involved *causal* connections between basic predicate tokenings and the properties they express, they would undoubtedly have agreed, unless they appealed to *immediate sensory acquaintance*, and thought that that cognitive-epistemic relation did not involve causation.) For these other expressions, Kripke brings in two further elements: an initial *baptism* or *dubbing* event, which *fixes* the

reference, and *causal-historical chains of transmission*, which *transmit* and *preserve* the reference.

c) Jackson explicitly insists that two-dimensionalists need not be descriptivists about reference generally. They can take Kripke's points on board—and in some sense, *must* take them on board. For they are *not* going to claim that all expressions have the sort of C-intensions that descriptions have. They are going to appeal to the descriptive A-intensions to fix reference, and thereby to fix C-intensions. And the idea is that they can think of the reference-fixing in each world as being settled by the descriptive A-intension and the way that world is, just on the model of a Kripkean using a description to fix a reference at the initial baptism.

d) But here, I think, we run into a problem. It is of the essence of Kripke's account of reference-fixing by descriptions (leaving aside ones that have demonstrative elements) that although baptizing by description—performing some such speech act as "I will call the next person to walk through the seminar door Nextone,"— makes *a priori* knowledge of contingent matters of fact (since the dubber knows *a priori* that Nextone is the next one through the door, but in some possible worlds, Anil—who, let us say, *is* the next one through the door, and so is in all possible worlds is the one we are referring to as 'Nextone'—is *not* the next one through the door, which is accordingly a contingent fact) possible *for the one performing the baptism*, what description was used to fix the reference is typically opaque to (and quite irrelevant to) ordinary users of the term, who are linked to the initial baptism only by a causal-historical chain [I think it is best thought of as an *anaphoric* chain, but never mind] of transmission. That is, for the sorts of expressions Kripke discusses, the reference-fixing descriptions associated with those expressions are not epistemically available to those using the expressions. Those descriptions *cannot* be A-intensions, for they fail the condition of *Semantic Transparency*. To revert to an earlier point, the *experts* connected with some natural-kind term may know how its reference was fixed, but the rest of us do not. This is one of the reasons Jackson explicitly excludes proper names from his discussion, and one of the reasons his account has trouble with technical natural-kind terms. He must think of the latter as resolvable into theories that can be expressed using simpler natural kinds, which *can* (he insists) be thought of as having their reference fixed by associated folk theories.

e) The point I want to make, however, is that Kripkean reference-fixing cannot be made sense of without liberal appeal to the concept of a <u>cause</u>. As he describes them, many dubbings fix the referent as what is *causing* a particular phenomenon, or is the *common cause* of a constellation of phenomena. This is true of natural-kind terms as well as proper names, and we have just seen that it is true as well even of secondary-quality predicates such as 'red', construed as response-dependent. This means that for Jackson to be entitled to say, as he does "A-intension is an *a priori* notion…because the A-extension in *w*…does not depend on which world in fact is actual…[I]n saying this we are saying what Kripke was saying when he said that sentences of the form 'K is …' where the ellipsis is filled in with whatever reference fixes to K are *a priori*," he must have a story about the A-intension of 'cause'. And as we have seen, that means it must either be a *basic* A-intension, or it must be specifiable as a Ramsified folk-theory expressible in

terms that *do* have basic A-intensions. Unless he can appeal to *causes* in A-intensions, the constraint of *Semantic Determination*—that the right references get fixed—cannot be satisfied on Kripkean grounds. But as we have seen, satisfying the *Epistemic Transparency* and *Dual Independence* constraints drives Jackson inexorably towards identifying *basic* A-intensions with *purely phenomenal* properties. It is, of course, a good question whether even these can meet the three constraints on A-intensions. But even supposing they can, descriptive theories expressed purely in phenomenal terms have no prospect (certainly not for Kripke) of meeting the constraint of *Semantic Determination*.

    f)

9) For all these reasons, Jackson is committed to walking a very fine line. Doing that requires him to endorse a very strong form of semantic reductionism. The A-intensions of *all* concepts must be expressible in terms of descriptions, Ramsified folk theories, couched ultimately in purely *phenomenal* terms. For those are the *only* terms it is remotely plausible could have the *epistemic* features of *basic* A-intensions.

10) [Finale: What are the A-intensions of *psychological* concepts and propositions, such as believing that George Washington is the father of this country? Jackson's physicalist must address *these* properties. But his apparatus gives us no idea how to associate A-intensions with them. In *this* context it is *not* legitimate simply to exclude proper names and demonstratives from consideration, for they contribute to the contents of the psychological states being attributed, and such attributions must turn out to supervene on the physical. (If the only intentional states that supervene on the physical turned out to be those whose contents could be specified without using proper names, demonstratives, or recalcitrant natural-kind terms, the physicalist's thesis would not have been borne out.)

11)

12)

13) Note: discuss *rigidification* of the fixed reference—what is made explicit by Kaplan's 'dthat' operator—and its importance for the transition from A-extension (and so, because they always coincide at *w* when we are "considering it as the actual world", the C-extension) to C-*intension* (from the perspective of *w*, as it were).

14) t