The A Priori Entailment Thesis

Commentary on Frank Jackson's *From Metaphysics to Ethics,*
Philosophy and Phenomenological Research, Vol. LXII, No. 3, May 2001.

Katalin Balog

Frank Jackson's new book *From Metaphysics to Ethics* is impressive for the grand and coherent vision it offers of the connection between metaphysics and conceptual analysis. The structure of the proposal is both clear and simple. Here are the main outlines:

a) One of the main tasks of metaphysics is to propose a kind of fact (or vocabulary) as fundamental and then locate other facts with respect to the fundamental facts. Physicalism, for example, is the claim that

PHYS: (T)(T -> □ (K –> T)

where K is the complete description of the world in the fundamental vocabulary of the true physical theory, including the fundamental laws of physics, and T is a sentential substitutional quantifier.[1]

b) Contra Quine, there are *a priori* truths that can be discovered by *conceptual analysis.* The method of conceptual analysis involves asking oneself whether one would apply a given concept in various situations.

c) Metaphysics and conceptual analysis are connected by the *A Priori Entailment Thesis (AE)*

AE) if physicalism is true, necessities of the form KεT are among the a priori analyticities (p 82).

This means that physicalists should be committed to the astonishing claim that anyone who grasps the concepts characterizing the fundamental physical facts would be in a position to figure out a priori the relationship between fundamental truths and truths of chemistry,

biology, psychology, ethics, a so on – provided she also grasps concepts of chemistry, biology, psychology, ethics, etc.

If AE is true then those who doubt that there are a priori entailments between physics and statements concerning a particular subject matter S will be forced to eliminativism, non-cognitivism, conceptual revision, or dualism with respect to S. And many philosophers do doubt that there are such entailments with respect to some of the most interesting matters discussed by philosophers (e.g., consciousness, thought, ethics). So it is imperative not only for Quineans who are skeptical about the analytic/synthetic distinction in the first place, but also for physicalists who accept the distinction but also think that qualitative, intentional or normative statements cannot be deduced a priori from the physical, to take a closer look at Jackson's arguments.

I) The two-dimensional framework

According to Kripke, certain concepts, for example, *water* have certain descriptions associated with them, e.g., "clear, odorless liquid" but are not equivalent in meaning to those descriptions. Rather, the description contingently fixes the reference of the expression. On Kripke's account, *water* is a rigid designator referring to H_2O in every world, *even though it may be that at some of those worlds H_2O is not a clear, odorless liquid and the clear, odorless liquid at those worlds is not H_2O.*

This suggests that we can associate with the concept *water* two distinct functions from worlds to references. One function picks out at each world the actual reference of the concept, that is, H_2O. The other picks out at a world **w** the kind (if there is one) that *water* would refer to *were **w** the actual world*. In Jackson's terminology, the first of these functions is called "C-intension", since it has to do with evaluating the semantical value of a concept in *counterfactual* worlds. The second he calls "A-intension", since it provides the semantical value of a concept in worlds *considered as actual*. Jackson suggests that A-intension is relevant to *grasping* a concept.

[1] This is strictly true only for "non-global" statements. A non-global statement is one whose truth value is decided by the fundamental facts and doesn't also require that these facts *are* the totality of fundamental facts.

Each thought expresses both an A-proposition and a C proposition. The A-proposition of a thought T assigns the truth-value 'true' to a *centered* world **<w,x>** if T would be true *were it thought at* **<*w,x*>.** E.g., according to Jackson, the thought *Water is wet* is true in a world *considered as actual*, iff the watery stuff in that world is wet. The C-proposition of a thought is just its ordinary content; the C-intension of the thought *water is wet* assigns the truth-value 'true' to a world **w** iff $H_2O$is wet in **w**. Some thoughts may express the same proposition as their A- and C-intension.

The two-dimensional framework is not controversial. It is based on the simple idea that meaning is a function of factors involving the thinker's mind and factors outside the thinker's mind. What is controversial is the nature and importance of A-intensions. Jackson thinks that the relevant internal aspect of meaning has to do with a concept's inferential role, which can be captured by descriptions associated with the concept. For example, *water* has an internal aspect involving narrowly determined descriptions, very roughly, the description *watery stuff*, i.e., *clear, odorless, etc...liquid around here that fills the oceans and lakes, etc.* which, given facts about the actual world, determines that its reference is H_2O. Accordingly, Jackson thinks that A-intensions specify what someone understands when they grasp a concept and that this understanding enables one to know *a priori* truths connecting the concept with other understood concepts.

In general, whether one thinks A-propositions are of importance depends on what, in one's view, determines A-propositions. Given our minimal definition of A-intension, A-propositions are determined by 1) how thoughts (and concepts generally) are individuated (what features are kept constant as we go from world to world, or rather - since A-intension is a function from *centered* worlds to truth-values - from context to context) and 2) how the C-intension is determined in a context. On some accounts of thought content A-propositions are completely unimportant. For example, on Fodor's theory,[2] concepts are individuated by syntax and orthography (or whatever corresponds to orthography in the language of thought) and by reference. Only the syntax and orthography is internal. Reference is externally determined by what the concept asymmetrically depends on at a world. Inferential role plays no role in individuating a concept although as part of the context at a world it may play a role

[2]Jerry Fodor: *Concepts: where cognitive science went wrong,* Oxford ; New York : Clarendon Press, 1998.

in partly determining asymmetric dependence. So on Fodor's theory, the A-intension of, for example, "cow" is a function which maps a world w onto whatever "cow" asymmetrically" depends on (if anything) at w. On this view, A-intension is not very interesting - it doesn't play a role in psychological explanation, doesn't underwrite analyticities, etc. On Holistic conceptual role theories, all of the conceptual role is held constant. Some such views treat reference as deflationary, others as causal and not determined by the conceptual role and others as what best fits the conceptual role. Each of these views have a different gloss on what A-intension (narrow content) comes to.

II) Jackson's two-dimensionalism

Jackson thinks that a thought's A-proposition is *a priori available* in that (assuming physicalism) knowing the full physical description would enable one to know the A-proposition's truth value. What underlies this ability is that being a competent user of one's concepts enables one to know a priori how contingent facts about the actual world figure in determining what one's concepts refer to. According to this proposal, given a full (supposedly fundamental) description of any possible world, one can figure out a priori whether, *were that the actual world*, there would be any water, trees, spiders, consciousness, etc. there. Let's call this the *A Priori Availability Thesis (AA)*. This tries to capture the idea that A-intension is not only *specifiable for the subject* from a third person point of view; it is also *a priori accessible to the subject* from the first person point of view. Jackson's answer to the question as to how A-intensions are determined can be reconstructed thus: 1) we take from world to world our concepts which are (narrow) descriptions revealed by conceptual analysis 2) in each (centered) world considered as actual they pick out whatever these narrow descriptions apply to. Moreover, and this is an important further gloss on the nature of our concepts, the descriptions associated with our concepts are such that they enable us, given a full fundamental description of a world, to tell what they would apply to *were that world the actual world*. This is quite an astonishing view that mere understanding should provide us with such amazing abilities.

There is no doubt that such a notion of A-intension provides us with a notion of content distinct from ordinary content (C-intension) that is both psychologically and

philosophically important. The question then becomes, do our concepts have A-intensions in Jackson's sense? Let's see a little more closely what is involved in proposing that they do.

1) For someone to be able to tell, given a full fundamental description of the world considered as actual, whether a thought would be true there, one would have to understand the fundamental language. But it is implausible that any one currently - even leading physicists - possesses the requisite fundamental concepts. AA, as it is stated, is an idealization; what Jackson has in mind is what an ideal logician, *once she has learned the fundamental language of physics,* and not being bound by time constraints, powers of physical endurance, etc. could figure out under optimal circumstances.

2) The fundamental description has to be narrow if A-intension is to be determined by matters inside the head. But this is not uncontroversial. On the Lewis-Ramsey account[3] of theoretical terms the theoretical terms of fundamental physics refer to that property- whatever it is- that actually satisfies a certain theoretical role. But it may well be that different properties satisfy a given role in different worlds.

3) What enables us to tell what our concepts refer to, or whether our thoughts are true in a possible world considered as actual, are presumably *descriptions* associated with our concepts, something like *the property X such that X plays the role R*. These descriptions are yielded by conceptual analysis. Again, the content of R has to be narrowly determined, given that A-intension is supposed to be a narrow matter. Moreover,

4) The descriptions comprising R have to provide enough connections with the fundamental language for the derivations to go through. Given physicalism, phenomenal concepts, e.g., can only enter these descriptions if they themselves are analyzable in ways that facilitate derivations from basic physical descriptions. Many physicalist would argue that phenomenal concepts are not so analyzable.

It seems to me that, given physicalism, the only hope to get all of the above is if each concept has a Ramsey definition whose only non-logical terms are the simple predicates of fundamental physics and if these predicates are rigid designators whose A and C intensions coincide. A concept's C-intension then would result from taking its extension to be whatever satisfies the Ramsey-definition at the actual world, and the same thing in counterfactual

[3]David Lewis: "How to define theoretical terms", *The Journal of Philosophy*, Vol. 67, No. 13. pp. 427-446.

worlds; it's A-intension picking out at a world whatever satisfies the Ramsey-definition at that world. The Ramsey definitions should employ only logical and mathematical concepts, certain topic neutral concepts like "cause", "chance" and the fundamental concepts (which themselves have to be narrow). This latter requirement is very peculiar since normally one would think of the non-logical concepts in a Ramsey definition as including common sense concepts (pointers, color concepts, phenomenal concepts, etc.), not concepts of fundamental physics. In fact Lewis introduced the idea of using Ramsey definitions (following Ramsey) as a way of defining theoretical concepts; i.e., those of fundamental physics.

The bottom line is that assumptions 1-4 are very controversial and one would need very strong reasons, preferably a very good argument to believe in them. Short of such an argument forthcoming it is more reasonable to believe that, even though there may be a matter of fact about what a term refers to in other contexts, it may be that one doesn't really know what it is by merely asking oneself what one would apply the term to. In the following section I will explore Jackson's argument for a prioricity and for his version of two-dimensionalism.

III) Jackson's argument for the a priori

On p. 53 Jackson takes up Quine's challenge and provides a brief sketch of an argument for the existence of a priori, analytic truths. This is the only formal argument he gives for the central doctrine that conceptual analysis has a priori results; so it will merit special attention. I first reconstruct the argument premise by premise; I try to follow Jackson's exposition here very closely. Subsequently I will interpret and discuss them one by one.

1) Telling how things are requires representation that somehow effects a partition in the possibilities.

2) Representations effect partitions in how things are, independently of how things actually are (independently of which world is the actual world).

3) There exist two representations, R_1 and R_2, such that the actual-world independent partition effected by R_1, and the actual-world independent partition effected by R_2 is such that the set associated with R_1 is a subset of the set associated with R_2.

4) If there are two representations, R_1 and R_2, such that the actual-world independent partition effected by R_1, and the actual-world independent partition effected by R_2 is such that the set associated with R_1 is a subset of the set associated with R_2, then we can know 'If R_1 then R_2' independently of knowing what the actual world is like.
5) What we can know independently of knowing what the actual world is like is a priori.
6) 'If R_1 then R_2' is a priori.

1 is the possible world account of propositions: telling (or thinking) how things are *via* expressing propositions involves effecting a partition in the possibilities, namely, between the set of possible worlds where the proposition is true, and the set of possible worlds where it is false. 1 merely provides background for the argument.

2 seems to be saying that if thoughts were to express propositions such that what proposition is expressed by a thought depended on the actual world (the context) in which it occurs, then they should also express propositions such that what proposition is expressed did not depend on the world in which it is thought. That is, unless thoughts had narrow content (as well as wide content in some cases), thoughts couldn't have content at all. This is the reverse of the externalist thesis according to which unless some thoughts had wide content no thoughts could have content at all.

In one sense, 2 is obvious. As we have seen, it is possible to define a notion of A-intension for concepts and thoughts, simply as a consequence of the fact that meaning is determined by factors "inside the head", on the one hand, and by factors "outside the head", on the other. A-intension is determined by taking the whole head, or some relevant part of it (a term in Mentalese, a conceptual role, etc.) from world to world and assigning to each world whatever the concept would refer to in that world *were that world the actual world.* The content thus determined is obviously narrow as it is determined entirely by "what's in the head". Let's call this weak interpretation of 2. The weak interpretation of 2 is only committed to the existence of *some* function from centered worlds to truth values that is determined by factors internal to the mind (or the head, if we are to remain in the physicalist framework). And as we have pointed out, this much is guaranteed by the trivial fact that content is determined by both factors "inside the head" and external to it. 2 then comes to this:

2*) Representations have A-intension in the basic sense outlined in the general framework of two-dimensionalism.

But it is unlikely that 2* is what Jackson has in mind; if all he head in mind is 2*, he wouldn't have to argue for it - it would merely be a simple point of logic. But he actually provides something like an argument for 2. Suppose, he suggests, that it is impossible to effect a partition among the possibilities independently of how things actually are, i.e., independently of which world is the actual world. This comes to the claim that thoughts don't have A-intensions. Then what partition our thought makes is always relative to the way things actually are. In one world one partition is made, in another world a different partition. This would mean, according to Jackson, that we could never say how things are; we could only say how things are if… "This is a very radical doctrine. It is not that we cannot say with complete precision how things are. We really cannot say how things are at all." (p.53)

It is not immediately clear what Jackson's point is. The claim might be that unless thoughts have A-intensions *of a certain specific nature*, we couldn't communicate, or know what we mean, in some important sense of knowing. In what follows I will try to make this precise: what exactly Jackson thinks A-intensions have to be like. And I will tackle this question from the point of view of what A-intensions have to be like for Jackson's argument for the existence of a priori knowledge to go through. I think this approach will yield interesting results in that it will help showing where the argument goes wrong.

Premise 2 is not strictly needed to get the conclusion of the argument. The key premises are premise 3 and 4; they, together with the rather un-contentious premise 5 about the nature of a prioricity, imply the conclusion of the argument, i.e., that there are truths that are knowable a priori. The argument is obviously valid. The only question concerns the truth of 3 and 4. The role of premise 2 is to focus attention on the question: What do A-intensions have to be like to make premise 3 and 4 true?

Unless we assume that A-intensions are descriptions, or at least are determined by concept-constituting inferential roles, we won't get premise 3. Let's consider, for example, Fodor's take on concepts, according to which concepts are mental terms in whose individuation inferential roles play no role. A-intensions are then determined by what those

syntactically individuated mental terms would pick out in different possible worlds considered as actual. Since mental terms can plausibly refer to anything depending on the context, for no R_1 and R_2 will it be the case that the A-intension of R_1 includes that of R_2 (except for logical relations which we supposedly preserved). Merely taking the entirety of a concept's inferential roles from world to world to determine A-intensions will not do either: there has to be a distinction between concept constituting and other roles. Only if concepts are individuated (at least partly) by concept constituting conceptual roles, will both 3 and 4 come out plausibly true: a friend of the a priori will want the implication from, say "It's a cat" to "It is an animal" to be knowable a priori, but not the implication from "It's a cat" to "It has fleas". If we took, e.g., the entire conceptual role from world to world, and not only the concept-constituting part of it, then "Cats have fleas" would come out a priori, which is absurd. So only if there is already a distinction between the concept constituting and other inferential roles will we get out sensible a priori truths. Premise 2 would then read:

2**) Representations have A-intensions that are partly determined by concept-constituting conceptual roles.

However, this means that a priori comes out only if a priori went in - in other words, the argument for the existence of a prioricity *presupposes* a prioricity, a bad case of circularity.

The situation is even worse for the argument if Jackson intends to use it to support the A Priori Entailment Thesis (AE). It is easy to see that conditionals of the form

K -> T

would only come out a priori if the A Priori Availability Thesis (AA) were true, since one could derive a priori any truth from the full physical description of the world only if it was possible to know, just by understanding a truth, whether, given the full fundamental description of a world, it would be true in that world. This is exactly AA. So the dialectical situation is this: to get AE, one needs to presuppose not only the existence of analytic truth in general, but more specifically, one needs to presuppose the truth of AA. But AA is an equally contentious claim: it puts very strong conditions on concept possession, as we have seen in Section 2. This leaves Jackson's main thesis, the A Priori Entailment Thesis, an article of faith that, on general grounds, there is not much reason to hold.