

=====
Distributed for La Pietra Conference June 21-23/2004
Paper is forthcoming (2005?) in volume of essays in honor of Frank Jackson edited by
Ian Ravenscroft to be published by OUP with replies by Jackson.
Sorry it is a bit long. Some may wish to skip Section 2 since it surveys the space of
familiar replies to the KA.
=====

Robert Van Gulick
Philosophy 541 HL
Syracuse University
Syracuse, NY 13244-1170
RNVANGUL@syr.edu

Jackson's Change of Mind:

Representationalism, *A Priorism* and the Knowledge Argument

1. Introduction

Few arguments in the recent philosophy of mind have generated as much discussion or controversy as Frank Jackson's Knowledge Argument, and none has had a more surprising history. Having offered the Knowledge Argument (KA) in 1983 and ably defended it for fifteen years against a wide variety of objections, Jackson did an about face in 1998 and disavowed it without really explaining his reasons for doing so. Only several years later in his paper "Mind and Illusion" did he offer a detailed account of his change of mind and explain where he had come to believe the argument goes wrong. His current rejection of the argument is essentially coupled with the acceptance of another controversial if widely held view, namely the Representational Theory of Consciousness or at least of Conscious Sensory Experience (RTS) which he regards as the sole means to defeat the KA and avoid the cognitive illusion that generates its intuitive appeal.

Jackson's current position thus involves at least three elements: a rejection of the Knowledge Argument, an acceptance of the Representational Theory of Conscious Sensory Experience, and a claim that the RTS provides the only effective grounds for rejecting the KA. My present concern is primarily with the third of these claims and with the question of whether one must accept the representationalist view in order to avoid the anti-physicalist conclusion of

the Knowledge Argument? I also have reservations about the representationalism, but I will focus here on its import for the knowledge argument rather than on the RTS itself. Determining whether it is the *sole means* to refute the KA in fact involves answering two component questions: Does the RTS provide *a* means to refute the KA? And are there any *other* ways to refute the KA that do not rely on the RTS? According to Jackson the respective answers are “yes” and “no”, but both are open to debate. Many plausible non-RTS objections to the KA have been offered over the past two decades, and some may in fact provide valid means of refuting it without needing to accept the RTS. Moreover, it is not obvious that the RTS in fact suffices to undermine the KA; adoptive defenders of the newly orphaned KA may still have resources to defend it against the line of criticism the RTS supposedly provides.

I have elsewhere earlier surveyed the space of critical objections to the Knowledge Argument. A decade ago (1993), I described four main lines of objection then found in the literature, and more recently (2004) I have added two others. None of the six relies in any obvious way upon the RTS, so I will begin by briefly reviewing both those old and new options in sections 2 and 3 respectively. In section 4 I will consider where, if anywhere, Jackson’s own present critical view fits in the logical geography of options. I will then turn in section 5 to the question of whether the RTS is in fact either necessary or sufficient for refuting the KA.

2. The Space of Existing Replies.

The Knowledge Argument appeals to Jackson’s now famous thought experiment of Mary the hypothetical super color scientist who supposedly knows all there is to know physically about having a red visual experience but who has spent her entire life in achromatic isolation and thus herself never had any such experience. The key intuition is that when Mary is finally released and has her first red experience she will learn (or come to know) something that she previously did not. Her epistemic gain is alleged to refute physicalism, which if it were true would have left nothing for Mary to have learned given her prior comprehensive physical knowledge. The master argument can be stated simply as

Argument KA.

P1 Mary before her release knows *everything physical* there is to know about seeing red.

P2 Mary before her release does not know *everything* there is to know about seeing red because she learns something about it on her release.

Therefore:

C1 There are some truths about seeing red that escape the physicalist story.

C2 Physicalism is false, and phenomenal properties cannot be explained as (or identified with) physical properties.

The argument is simple but bold. It asserts a major metaphysical conclusion with sweeping empirical consequences on the basis of intuitions about a purely hypothetical case. As such it shoulders a heavy burden of proof. The plausibility of the argument and of any assumptions on which it relies must be weighed against the countervailing force of widely held and apparently well confirmed theory.

The Knowledge Argument is also an example of what I have elsewhere (2003) labeled a “boomerang argument”. The characteristic feature of such arguments is that they reach over into the epistemic/cognitive/conceptual domain to a supposed fact about our representation of reality and then use it to swing back to reach a conclusion in the metaphysical/ontological realm of facts about reality itself. More specifically, boomerangs often move from supposed gaps or lacks in our representations or concepts of the world to conclusions about objective gaps within the world itself and ontological distinctions among the real things in it. This is clearly the case with the Knowledge Argument as well as with dualist arguments based on the alleged logical possibility of zombies (i.e. beings physically indistinguishable from normal humans but devoid of any conscious mentality).

Boomerang arguments have a long history that goes back at least as far as the early seventeenth century to Descartes’ argument for mind/body dualism by appeal to the supposed conceivability of the disembodied persistence of his mind. They have an equally long history of

criticism, and they seem especially vulnerable to charges of illicitly inflating a subjective or conceptual gap into an objective or ontological one. Nonetheless, each argument deserves consideration on its specific merits; perhaps some boomerangs can validly be thrown and soundly find their target. However, they also merit special scrutiny, and the Knowledge Argument perhaps even more than others since it relies on a doubly epistemic route to reach its result. It invokes intuitions about what Mary would and would not know in highly non-actual conditions, and then uses those supposed epistemic lacks to support its anti-physicalist conclusion. To push the boomerang metaphor a bit, the argument has to circle twice within the epistemic realm before landing back on the ontological metaphysical side. Perhaps the Knowledge Argument can fly the course, but it is open to multiple lines of attack by those who believe it can not.

As I have shown earlier(1993, 2003), most of the critical replies in the literature can be sorted by their answers to a series of diagnostic questions that identify the points at which they take issue with the argument's line of reasoning. Those questions can be used to generate a tree diagram (figure 1) that locates the various critical options in logical space. Some challenge the argument early on; others go some distance with it before disagreeing.

Q1. Does Mary learn anything or gain knowledge when she first experiences red?

//	\\
<u>NO (Churchland, Hardin)</u>	<u>YES</u>
X Physicalism not refuted.	
	∨

Q2. What sort of knowledge does Mary gain? Is it strictly *know-how*, or does it include new *knowledge of facts or propositions* .

//	\\
Gains only new know-how.	Gains new knowledge of propositions.
<u>(Lewis, Nemirow)</u>	
X Physicalism not refuted.	
	∨

Q3. Does Mary upon her release come to know *new facts* and new *propositions* ?

//	\\
NO. She only comes to know old propositions in a new way.	YES. She comes to know new propositions.
<u>(Tye, Horgan, Churchland)</u>	
X Physicalism not refuted.	
	∨

Q4. On what mode of individuation does Mary learn a new proposition?

//	\\
Only on a fine grained mode.	On a coarse grained mode.
<u>(Lycan, Loar)</u>	<u>(Jackson pre-1998)</u>
X Physicalism not refuted.	Physicalism refuted.

(Each “X” marks a point at which the knowledge argument is blocked.)
Map of Four Critical Options

Figure 1

The first diagnostic question asks whether Mary would in fact learn anything or gain any knowledge when she first experiences red. The argument assumes and requires that she would, but some physicalists have held otherwise and challenged the reliability of our intuitions about such highly counterfactual cases. They note that by stipulation, Mary knows *everything physical* there is to know about having a red experience. Such comprehensive physical knowledge is so far beyond what we possess today, indeed beyond what a real human being with our inherent cognitive limits could possess, that we should be cautious in evaluating our intuitions about such cases. There is the danger that our intuitions might fail to reflect the absolutely unlimited nature of Mary's stipulated physical knowledge, and that we might simply imagine someone who knows a lot more of the sort of physical knowledge we possess today but still falls far short of totally comprehensive physical knowledge. Given how unlike our current physical knowledge Mary's state would have to be, one might doubt that we can have clear or reliable intuitions about what she would or would not understand from such a physically comprehensive epistemic perspective. Perhaps Mary, or super Mary, when released would exclaim, "Yes it is just as I knew it would be." Perhaps not, but there seems reason for caution in making any such predictions, given how far from epistemic reality our intuitions are asked to stretch.

If one concedes that post-release Mary gains at least some knowledge, the next question in the decision tree Q2 asks whether she gains any new propositional knowledge or only new know-how. According to the so called "ability reply" Mary gains only new epistemic abilities or skills to recognize, imagine, and perhaps compare red experiences; she gains no new propositional knowledge. If we think about a gain of propositional knowledge, as David Lewis did, as a reduction of uncertainty, then the ability reply denies that post-release Mary eliminates any prior open epistemic possibilities. Her having experienced red does not enable her to locate herself more narrowly or precisely in logical space; her comprehensive physical knowledge already sufficed to identify her location. The differences and enhancements in her epistemic situation are solely matters of acquiring new skills, but these newly acquired abilities to recognize or imagine red experiences supposedly do not constitute the acquisition of any new

propositional knowledge. The epistemic additions are all just matters of new practical know-how and not of any new knowledge that. Mary no doubt acquires *some* important new know-how, but whether that is *all* she acquires is less obvious and more controversial. Lewis and Nemirow see only new abilities, but others including Jackson in the past, insist there is more to Mary's gain. (Jackson now sides with Lewis and Nemirow, though he claims to have followed a different route to arrive there.)

Those who believe Mary gains more than mere know-how, move on to Q3, which initially may sound the same as Q2, though it is importantly different. Q3 asks whether Mary gains any knowledge of new propositions. The difference between Q2 and Q3 concerns the scope of "new". Q2 ask about new propositional *knowledge*, whereas Q3 asks about knowledge of new *propositions*. Mary might gain the first but not the second if her propositional knowledge could change in ways sufficient to count it as new knowledge even if it does not involve an extension in the range of propositions she knows. Could there be sufficient changes in the knowledge states themselves, even if she gains no new objects of knowledge? We might for example count Mary as having new propositional knowledge if she gained a new way of conceptualizing, accessing or representing to herself some propositions that she already new. The proposition that she knows may be the same, but if her *way* of knowing it is sufficiently different we may want to count it as a different epistemic state and as a gain in propositional knowledge.

Some have disparaged any such gain as trivial. David Lewis argued that if learning a new way of representing a previously known proposition sufficed for a gain in propositional knowledge, one could gain lots of new new knowledge about non-linguistic matters just by learning a new language. If I learn Urdu (his example) and gain the ability to represent the proposition that snow is white (" " } I do not increase my propositional knowledge about meteorological matters (as opposed to what I may learn about Urdu itself). Lewis seems right about the example, but its relevance to the Knowledge Argument is less clear. The representational or conceptual changes involved in Mary's case may not be analogous. The

modes of cognitive access and representation made available to post-release Mary likely differ from those afforded by neuroscience far more than Urdu does from English as a means of cognizing meteorological propositions.

The distinction to which Q3 points is sometimes couched in terms of modes of presentation; one might know a proposition under (or through) one mode of presentation but not under another. I might know the glass contains water under a perceptual and folk conceptual mode of presentation, but not have access to the same proposition under the a more theoretical and abstract chemical mode that presents it in terms of H₂O molecules. Gaining the latter mode of access seems an enhancement of my overall epistemic status even if the proposition and associated property in the latter case is the same as that involved in the former (being water just is the property of being composed of H₂O molecules). Similarly the modes of presentation and access that Mary gains through her experiential, introspective and empathetic systems when she sees her first tomato would seem to enhance her cognitive and representational access to propositions previously known to her only through theoretical neurobiological concepts.

Thus some critics of the knowledge argument have aimed to block its progress at Q3. They concede Mary gains new propositional knowledge, but only in the sense that she gains significantly new modes of representing, conceptualizing or cognitively accessing propositions that she already knew. Moreover, they argue that such a gain is not inconsistent with physicalism and provides no reason to reject it, except perhaps in extreme reductionist forms that treat physical theories as in themselves sufficient for the practical representation of everything real. Most physicalists are far more pluralistic about the diversity of modes of access and presentation that we need. Few, if any, would advocate doing economics with the resources of physical theory. Nor should we require neurobiology to duplicate or provide all the useful modes of cognitive access and engagement available through the first person perspective. Mary's apparent epistemic gain might thus be fully explained in terms of her gaining new modes of access and presentation to facts and propositions she previously knew through the means of physical science.

Defenders of the KA may deny the strategy can be used when it is facts about modes of presentation themselves that are centrally at issue in the argument. It is what Mary learns about the experiential mode of sensing red that supposedly escaped her prior physical knowledge. Physicalists might counter in return that Mary's prior knowledge did indeed include those very propositions about experience, though her access to them was through non-experiential modes or means. As long as she learns no new propositions, physicalism is not in jeopardy.

Others go even further with the argument before making a late swerve away from its anti-physicalist conclusion. They concede not only that Mary gains new propositional knowledge but also that she comes to know propositions she did not previously know, and yet they deny that her gain refutes physicalism. The crucial issue is how one individuates propositions and thus what counts as coming to know a new one. Q4 asks on what mode of individuation does Mary count as learning a new proposition? There many possible schemes of individuation one might use, some more coarse grained or fine grained than others. The latter distinguish among propositions that are regarded as the same by the former.

In a way Q3 and Q4 might be seen as two ways of asking the same question, or two sides of the same critical coin. Where does one draw the line between two ways of knowing the same proposition and knowing two distinct but closely related propositions? Is coming to know that the glass contains H₂O coming to know a new fine grained proposition or merely gaining a new mode of access to a previously known coarse grained one? The fine grained mode distinguishes propositions in terms of conceptual or representation features that more coarse grained schemes disregard as external matters about modes of access rather than the identity of propositions themselves. There may be no "correct" place to draw the line, and its exact location may be of no real importance to the Knowledge Argument. Physicalist critics might try to block it by explaining Mary's total epistemic change as a matter either of her coming to know a old proposition in a new way or of gaining knowledge of a new proposition in the fine grained sense.

What the physicalist critic seemingly cannot concede is that Mary learns a new proposition on a coarse grained mode of individuation. Her doing that would seem to entail the

existence of properties cognitively inaccessible to her, rather than merely concepts or modes of presentation that she could not grasp. And if there were such properties of which she lacked prior knowledge then it would also seem to follow that there are possibilities and uncertainties she could not eliminate until she had had a red experience, possibilities defined in terms of the presence or absence of the relevant property(ies). Since by stipulation she knew about *all the physical properties* beforehand, it would thus seem to follow that there must be at least some non-physical properties, and physicalism would be refuted. Thus with respect to Q3 and Q4, the key issue dividing supporters and critics of KA seems to be whether or not Mary comes to know a new coarse grained proposition.

3. Two Newer Critical Options.

Or so I once believed and wrote (1993). However, more recently (2003) I have come to believe there are at least a couple of further critical options, including one that may allow the critic of the KA to concede even that Mary learns a new coarse grained proposition while still denying that her doing so refutes physicalism. The two new replies to the KA more or less bracket the earlier four and can be located on an expanded tree diagram (figure 2) that adds one question before the original Q1-Q4 sequence and another after, call them Q0 and Q5 respectively.

Q0 Is it possible for Mary prior to release to know all the physical facts and propositions about seeing red?

	//	\\
NO, not if there physical <u>subjective facts.</u>	YES	
X Physicalism not refuted.		

Q1. Does Mary learn anything or gain knowledge when she first experiences red?

	//	\\
<u>NO (Churchland, Hardin)</u>		YES
X Physicalism not refuted		

Q2. What sort of knowledge does Mary gain? Is it strictly *know-how*. or does it include new *knowledge of facts or propositions*.

	//	\\
Gains only new KNOW-HOW. <u>(Lewis, Nemirow)</u>		Gains new knowledge of PROPOSITIONS.
X Physicalism not refuted.		

Q3. Does Mary upon her release come to know *new facts* and new *propositions* ?

	//	\\
NO. She only comes to know propositions in a new way <u>(Tye, Horgan, Churchland)</u>	YES. She comes to know new propositions.	old
X Physicalism not refuted.		

Q4. On what mode of individuation does Mary learn a new proposition?

	//	\\
Only on a FINE grained mode. <u>(Lycan, Loar)</u>		On a COARSE grained mode.
X Physicalism not refuted.		

Q5. Would Mary's learning a new coarse grained proposition or her coming to better understand her location in logical space refute physicalism?

	//	\\
<u>NO (Nonreductive Physicalist)</u>		YES (Jackson pre-1998)
X Physicalism not refuted.		Physicalism refuted

(Each "X" marks a point at which the knowledge argument is blocked.)
Expanded Map of Critical Options

Figure 2

Although Q0 and Q5 come at the opposite of ends of the diagnostic chain, they turn on a common issue, namely whether or not physicalism can accommodate subjective facts and propositions. Facts are subjective in the relevant epistemic sense (Nagel 1974) if they are fully knowable and understandable only from a particular point of view, that associated with a particular form of experience. It is in that sense that Nagel claimed that facts about what it is like to be a bat are subjective. However, the existence of such facts or propositions refutes physicalism only if one assumes that *all* physical facts and propositions are objective and capable of being fully understood and known from many different perspectives and by many types of knowers with diverse experiential profiles. Strictly physical facts of the sort represented by physical theory do indeed seem knowable without restriction to any particular experiential point of view, but it is not obvious that all physical facts are of that sort, especially if one defines the domain of the physical to include all facts that are realized in virtue of underlying strictly physical facts.

If first person facts about experience are physically realized facts, then there may be some facts and propositions that are both physical and subjective. Moreover, some physicalists may be able to give a plausible account of how that might be. If one is a nonreductive physicalist and takes a pragmatic view of understanding and cognition, then it will hardly be surprising that there are physically realized facts that one cannot understand by using physical theory. Ontological realization does not entail that the methods sufficient for modeling the entities and actions at the level of underlying realization must suffice as well for modeling or understanding all the higher level aspects of the realized systems and their modes of interaction, especially since the relevant modeling needs to be done by particularly structured cognitive agents engaged and situated in a diverse range of contexts of application and interaction. Given the multiple parameters that condition those contexts of understanding and the intentional relations they subserve, it becomes all the less plausible to regard physical theory as a universal tool of understanding or as an all purpose medium providing commensurable translations of everything that might be validly represented or expressed by any other engaged and situated

cognitive system about any aspect of the physically realized world.

This is especially so when the system being understood and the one doing the understanding are one and the same, as they are in consciously self-understanding systems, such as a human or a bat. The nonreductive physicalist should not be at all surprised that such systems are able to engage their own physically realized dynamic features in ways that can not be duplicated by any equivalent external mode of cognitive access. If conscious minds involve a great deal of reflexive self-understanding, as I believe they do, there may well have aspects or properties that can not be fully understood from an outside perspective. Physicalists of a nonreductive and pragmatic bent would not suppose that third person physical theory must provide us with a means of grasping and understanding everything that physically realized self-understanding systems are able to understand about themselves. A commitment to universal physical realization carries no commitment to physical theory as a universal representational tool adequate for understanding everything real.

Indeed I have long argued (1985, 1992) that the existence of subjective facts is not only compatible with physicalism but predicted and entailed by the sort of teleo-pragmatic physicalism that I take to be most plausible. I will not repeat those earlier arguments in detail here, but give just the main idea. According to the teleo-pragmatic view understanding is always a matter of the potential for successful practical engagement, which often involves mutual reciprocal interaction and perhaps what might be called “causal resonance” or “causal harmonic engagement.” The intentional profile or content associated with such understanding is very much a function of the causal profile of that engagement and the sort of access it affords the understander to successfully interact with that which it understands. It is highly unlikely that we could fashion any structure through the use of external third person physical theory that would allow us to come close to duplicating the causal profile associated with the way in which a complex physically realized self-understanding system, such as a bat or human mind/brain, understands itself.

Nor need the teleo-pragmatic physicalist or any nonreductive physicalist, think otherwise.

The contexts of application for the respective representations are so dissimilar that it is unlikely we can use the one to replicate the causal interactive profile of the other to a degree that would make them intentionally equivalent. That at least is the main idea, and the details can be found elsewhere by those who may be interested (Van Gulick 1985, 1992). For present purposes, it should suffice to make at least a *prima facie* case for the possibility that subjective facts might coexist consistently with some plausible versions of physicalism.

That possibility provides us with an additional option for parting company with the knowledge argument, indeed with an option for “getting off the train” even earlier by disputing premise P1. P1 is generally conceded by the argument’s critics since it functions as a stipulation of the conditions about which we are invited to engage in hypothetical reasoning. Critics may note that no real human could in practice command as much as Mary is imagined to do, but the limits would seem to be merely contingent, such as those on memory or attention capacity. Thus there might seem to be no reason why someone could not at least *in principle* know all the physical facts as P1 supposes Mary does before her release. However, if some physical facts are subjective (in the sense of being understandable only from a specific experiential perspective), then Mary could not even in principle know them given the stipulated restrictions on her past range of experience.

Q0 asks, “Is it possible for Mary prior to release to know all the physical facts and propositions about seeing red?” In reply, a physicalist who believes in physically realized subjective facts should answer “No”, and thereby also reject P1 of the KA. Despite its status as a mere stipulation, one might refuse to concede it on the grounds that it embodies a hidden contradiction or a conjunction of conditions that can not be jointly satisfied. Mary is supposed to have never experienced red and yet to know all the physical facts about doing so. Perhaps she could in principle know all the “physical science facts.” However, if there are subjective physical facts, i.e. facts that are physically realized but cognitively accessible only from the experiential perspective of a certain range of physically realized self-understanding systems (whether bats or normal color experiencing humans), then Mary could NOT know all such

physical facts about seeing red given her stipulated lack of past experience.

Thus a physicalist critic who believes in the existence of subjective physical facts, as I believe she should, has good grounds for not taking even the argument's first step.

Let us turn then to our other new option and the final question in our expanded tree. Q5 asks, "Would Mary's learning a new coarse grained proposition or her coming to better understand her location in logical space refute physicalism?" I had earlier believed that the answer to Q5 would have to be "Yes", but I now think there are plausible grounds for saying "No", at least if one unpacks the mental-physical relation nonreductively in terms of realization. I have given a fairly complete explanation of those grounds recently elsewhere (2004), but I hope a quick reprise of that discussion here will suffice to indicate the main idea.

One needs to consider two crucial and interdependent issues at the outset. First in what sense are higher level facts and propositions about realized mental properties new and different from those that concern their underlying physical realizations? Second, put in epistemic terms directly relevant to the Knowledge Argument, would knowing all the facts or true propositions that hold at the microphysical level suffice for knowing all the true facts and propositions that are realized at every level by those underlying microphysical facts? In what sense would knowing the first set settle every question about the latter? Does knowing all the realizer facts guarantee that one also knows all those that are realized? And would knowing all the microphysical facts leave any room to reduce one's semantic uncertainty about one's location in the space of possibilities?

If one accepts a principle of micro-physical supervenience, then it may seem that no such room could be left. If all the facts supervene on the micro-physical facts then once the micro-physical facts of the actual world are set, then so too are all the other facts. There are no subregions left for further information to eliminate. In particular the space specified by the micro-physical facts does not divide into distinct subregions that differ from each other with respect to any real non-physical facts. That is just what supervenience entails: no real difference without a micro-physical difference. However, one must take care in moving from facts about

ontology to matters of epistemology.

If coming to know a new coarse grained proposition requires locating oneself more narrowly in the logical space of worlds, then being able to locate oneself microphysically would seem to maximize the definiteness of one's location and thus to preclude learning any new coarse propositions. However, that implication depends crucially on what counts as knowing one's location in logical space. One may be able to specify a given world as actual and locate it in microphysical space, but still not have a full understanding of the structure of the overall logical space within which it is embedded. One might be like the lost visitor in a cartoon joke who confronts an informational display which announces "You are here *" but provides no further map to explain where "here*" is. Given supervenience, the microphysical facts may fix its location in that space, but in so far as one does not understand the structure of the containing space, one's knowledge of one's location is incomplete and may leave room for more than notional additions. And if one's cognitive mastery of the structure of the space is incomplete, so too may be the range of propositions one can know or even understand.

In what sense is the complex proposition that specifies the total realization base RB for some actual world object X a different proposition from that that expresses X's having the realized higher level property HP? If we think of propositions in the relatively coarse sense as functions from worlds to truth values, it seems that different functions would be associated with propositions that respectively specify objects as having RB or HP. On the assumption that HP like most higher level properties is open to a diversity of multiple realizations, the two functions will diverge in cases where objects realize HP in virtue of some base other than RB. The function for the HP proposition will assign T in such cases, but that for the RB will assign F. Given supervenience, everything which gets a T from the RB function will also get one from HP, but the converse does not hold because of multiple realization. Thus even on a coarse mode of individuation, the proposition that X has RB is not the same as the proposition that X has HP.

However, their nonidentity does not by itself guarantee the possibility that Mary could know the first but not the second. Even if the propositions are distinct, it still may be that

knowing the first in some way suffices for knowing the latter: either because knowing the realizer facts in general suffices for knowing the realized facts, or because it does so for particular reasons in the Mary case. Supervenience might seem to support such an entailment. As a matter of metaphysics or ontology, X 's having RB suffices for X's having HP (or at least does so *modulo* the natural laws of the actual world). So knowing the realizer facts and the relevant laws might seem to guarantee that one should also know the necessarily realized facts as well. One need only draw the consequences.

Of course, we often fail to explicitly recognize many of the less obvious consequences of our beliefs; mathematics otherwise would be much easier than it is. Moreover, Mary's lack is worse than that in the mathematical case. Her problem is not that of determining which of a given set of propositions are true (or follow as true), but rather of being able to even grasp or understand them. Mary's ability to understand all the propositions about the physical base need not seem suffice to guarantee her mastery of those about the realized properties, nor is there any guarantee that the concepts or modes of cognitive access needed to grasp propositions about higher-order realized facts are definable in terms of those that suffice for grasping propositions about the realization base. Indeed according to the nonreductive physicalist, we should commonly expect such failures of definability or translatability between the representational systems of the special sciences and those of underlying physical theory. They follow naturally from the pragmatic nature of representation and the way in which representational content depends upon the causal structures of both the cognitive agent and the object as well as on the larger dynamic context of engagement within which those representations get applied. There is little reason to suppose that we can fashion cognitively equivalent higher level tools out of the resources that suffice for successful engagement at the realization level. As noted above, physics just does not give us the conceptual and representational tools we need for doing economics or introspective psychology.

Returning to Mary, we can now see why pre-release she might lack a full understanding of propositions about seeing red, even if such experiences are just higher-order physically

realized states. Moreover, we can better appreciate the respect in which she might fail to understand the structure of logical space and thus of her actual location in it. In particular she may fail to adequately understand the structure of similarity relations that hold between the various situations and worlds in which the higher-order property of seeing red is realized by diverse physical substrates. In so far as Mary can not articulate those relations, her command of the structure of logical space may be inadequate to grasp either the nature of the realized higher-order property or the coarse grained proposition defined in terms of its presence or absence. If we treat coarse grained propositions as functions from worlds to truth values, Mary may lack cognitive command of the relevant function that attributes the higher-order property to particular experiencers. Thus prior to her release there may be coarse-grained propositions concerning physically realized properties that Mary cannot grasp, and in so far as she lacks an adequate cognitive command of the structure of logical space she may fail to fully appreciate her location in it. The physicalist thus might put off his disagreement with the KA until Q5, conceding in response to Q4 that Mary learns a new coarse grained proposition but still denying that her gain refutes physicalism.

Thus there are at least six points at which the physicalist might take issue with the KA:

- If she accepts the existence of physically realized subjective facts, she might reject P1 on the grounds that the stipulated conditions of the hypothetical contain an implicit contradiction. (Q0)
- She might deny that Mary learns anything new upon her release. (Q1)
- She might claim that Mary gains only new abilities and know-how. (Q2)
- She might concede that Mary gains new propositional knowledge but only in the sense of coming to know a previously known proposition in a new way. (Q3)
- She might concede that Mary comes to know new propositions but only on a fine grained mode of individuation. (Q4)
- She might allow that Mary learns a new coarse grained proposition but still deny that her doing so refutes physicalism, if physicalism allows for physically realized subjective facts. If it does, Mary's gain might be consistently explained as her coming to understand such facts and to

appreciate the structure of logical space in the ways needed to grasp the relevant propositions about them. (Q5)

4. Locating Jackson's Rejection.

Having surveyed the above six options, we can ask where in our tree, Jackson's own current objections to the argument might best be located. He places himself with Lewis and Nemirow as embracing the Ability Reply and accounting for Mary's epistemic gains as solely a matter of acquiring new abilities to recognize and imagine red experiences. Thus Jackson is most aptly placed as parting with the KA at Q2. However, there is a respect in which his current view also coincides sympathetically, if not quite logically, with those like Churchland who challenge the KA one step earlier by denying Mary would learn anything upon release. He takes a strong view about the reach of physical theory and claims Mary could, at least in principle, deduce *a priori* all the truths about seeing red from her comprehensive knowledge of the micro-physical facts. He believes she could in principle infer every true proposition about the experience of seeing red from the physical theory facts. Thus his current position might best be described as the conjunction of two claims:

- Pre-release Mary can deduce all the mental facts or propositions about experiencing red.
- Mary improves her epistemic status when she herself first experiences red, but only by gaining some new know-how and practical cognitive abilities.

Jackson also presents his view as crucially dependent on the representational theory of sensory experience (RTS), but how so? It is invoked to support the claim that Mary can deduce all the mental facts from the micro-physical facts. According to the RTS, experiences of seeing red have no mental properties other than their representational or intentional properties, and Jackson believes all the relevant representational facts are in principle deducible from the micro-physical facts. Although he does not endorse any specific account of representational content, Jackson asserts that the extant theories - whether causal, covariational or teleological - allow the *a priori* deducibility of all the facts about what is represented from a sufficiently complete

description of the micro-physical facts. If representational intentional facts are fully deducible from micro-physical facts, and sensory experiences such as seeing red have no mental properties beyond their representational intentional properties, then it would follow immediately that one can deduce all the mental facts and propositions about seeing red from the micro-physical facts. Since Mary supposedly knows all the latter, it would follow that she can know all the mental facts about red experiences before ever having one herself. An important qualification is immediately required. Jackson does not claim that Mary could *in practice* carry out any such deduction; it may well be beyond her actual calculational or reasoning powers, indeed beyond those of any actual human. Jackson claims only that it is *in principle* possible for Mary to carry out such an *a priori* deduction.

The role of the RTS is to deflect or rebut a putative objection to the claim of *a priori* deducibility. If the RTS were false, then conscious sensory states might have further properties whose nature or presence could not be deduced from the micro-physical facts. In particular, conscious sensory states might be thought to have phenomenal properties that were inaccessible through the deductive route. Indeed it is the belief in the existence of such properties that constitutes the “illusion” in Jackson’s title “Mind and Illusion.” If one believed, as Jackson apparently did in the past, that sensory experiences themselves had phenomenal color properties (or involved acquaintance with inner items having them), then Mary might not be able to deduce all the relevant propositions about such experiences from the micro-physical facts. If phenomenal redness (PHRED-ness) were an intrinsic property of red experiences (or of red sense data that were accessible only through such experiences), then there would be true propositions about experiences (or inner objects of experience) being PHRED or having PHRED-ness. Mary’s prior propositional knowledge would be incomplete unless she could deduce the required propositions about the presence or absence of PHRED-ness, but her mastery of the micro-physical facts need not supply her with adequate cognitive access to PHRED-ness or its nature. Thus one who believed in the existence of phenomenal properties would likely deny that Mary could *a priori* deduce all the propositions true of red experience, and argue

instead that Mary upon release would gain knowledge of propositions about phenomenal properties, such as PHRED-ness, that were previously beyond her ken. (Though we will see in section 5 that some physicalists might accept the existence of phenomenal properties but still contend that Mary could nonetheless deduce all the facts about experience.)

The primary dialectical role of the RTS in Jackson's current thinking seems to be to block the phenomenal properties objection to the deducibility thesis. According to the representationalist, there are no phenomenal properties such as PHRED-ness, so there are can be no propositions about them that pre-release Mary fails to know or is unable to deduce. On the representationalist theory, having a red experience is not a matter of having an experience with the property of PHRED-ness nor of being acquainted with a phenomenal object or sense datum with the property of PHRED-ness. It is simply a matter of being in a state that phenomenally represents external objects as being red. The only color property involved is ordinary RED-ness, and it occurs as part of the representational or intentional content of the experiential state not as a property of the state or any inner objects. If we use square brackets to indicate contents, the way to think about Mary's seeing a red tomato according to the RTS is roughly as

Mary visually represents [tomato is red].

The only occurrence of "red" is internal to the intentional content, and the relevant property being referred to is RED-ness a property of ordinary external objects. There is no reference to any phenomenal properties such as PHRED-ness, nor to any color properties of experiences or inner mental objects. Indeed the representationalist denies the existence of PHRED-ness.

If there is no such property as PHRED-ness, then are no coarse grained properties concerning which items have it or lack it. Jackson's appears to believe that if to the contrary he were to reject the RTS, he would be committed to existence of phenomenal properties such as PHRED-ness. And if so, then there would be propositions in the coarse grained sense that escaped Mary's prior knowledge, namely those concerning which items had and lacked the relevant phenomenal properties. Mary's pre-release knowledge would thus be incomplete in the strongly propositional sense that there were real possibilities and uncertainties she could not

eliminate despite her complete micro-physical knowledge. Since Jackson believes any such result would refute physicalism, he regards the RTS with its denial of phenomenal properties as essential for blocking the Knowledge Argument.

5. The RTS and the KA

As a matter of intellectual auto-biography, we have no reason to doubt the role played by the RTS in Jackson's change of mind about the KA. However, as a matter of logic, it is less clear that the RTS is either necessary or sufficient to undermine the Knowledge Argument. Though Jackson justifies drawing the line where he does by appeal to the RTS, some critics disagree with the KA at that same location in the question tree but without obviously relying on the RTS to do so. Others challenge the argument at different locations, none of which seem to rely on the RTS. Moreover, looking a bit deeper at the interplay between the KA and the RTS may give us a clearer understanding of just where the original argument may go wrong.

At least some physicalists should be happy to have other options for blocking the KA. Though the RTS is currently quite popular and ably defended, it is far from universally accepted, and those who oppose it can marshal some plausible concerns against it, especially against its claim that conscious sensory states have no mental properties other than their intentional representational ones. Inverted spectrum cases, as well as Ned Block's inverted Earth cases seem to indicate otherwise. Representationalists have of course replied and claim to have shown that such alleged possibilities pose no threat to their position. As noted at the onset, my current aim is not to adjudicate those disputes nor offer a bottom line judgment on the truth or adequacy of the representationalist view, but only to address its relation to the Knowledge Argument. In that respect, many physicalists who are less than fully convinced about the truth of representationalism, would probably prefer not to have their refutation of the KA held hostage to the fate of the RTS. Thus it is worth getting clear on the link between the KA and the RTS, even if one can not reach consensus about the RTS itself.

Though it may sound odd, I think the best way to see if the RTS is necessary for refuting

the KA is to look first look at the complementary question of its sufficiency. Doing so should help to clarify in just what way the RTS supposedly undermines the reasoning of the Mary argument. The RTS is of course not by itself sufficient in the strict sense to refute the KA. One must invoke at least some other claims, but Jackson seems to think the other assumptions one needs are all relatively obvious and not open to real challenge. However, some of them may be more problematic than he supposes, particularly the strongly *a priorist* assumption he relies on to conclude that Mary could deduce all the representational facts from the micro-physical facts.

As noted in the previous section, a key conditional in his reasoning has a conjunctive antecedent: *If* all the representational intentional facts are fully deducible from micro-physical facts, *and* sensory experiences such as seeing red have no mental properties beyond their representational intentional properties, *then* it would follow immediately that one can deduce all the mental facts and propositions about seeing red from the micro-physical facts. The RTS asserts the second conjunct, but one needs independent grounds for the first. What guarantees the *a priori* deducibility of the representational facts? Another way of putting the need is in terms of the dialectical role of the RTS in Jackson's rejection of the argument. As also noted in the prior section, the RTS serves to parry a objection that might raised by supporters of the argument concerning the non-deducibility of propositions about phenomenal properties. The RTS aims to moot that objection by denying the existence of such properties and thus of any facts or propositions about them. But even if that works, it is primarily a defensive maneuver and by itself does not establish the positive truth of the deducibility claim, which requires a prior logical basis.

Jackson appears to rely on two sources of support for the deducibility thesis, one more specifically concerned with the nature of intentionality and the other of a more general metaphysical epistemological sort. On the specific side, he appeals to the kinds of naturalistic theories found in the current literature, such as those that ground and determine matters of intentional content on facts of covariance, causal linkage or teleological function. Jackson believes that the analyses offered by such theories invoke conditions of the sort that could be

deduced (at least *in principle*) from the micro-physical facts. Thus such naturalistic analyses would provide an inferential bridge to deduce intentional facts from micro-physical ones.

There are several potential problems. First it is not obvious that all the determining facts could be deduced from the micro-physical facts. Causal and covariational facts might be seem safely deducible, but the case for deducing teleological facts is less obvious. Moreover, teleology may infect causal or covariational models as well, in so far as teleological considerations may partly determine which covariance or causal relations are relevant to content.

Secondly, the naturalistic theories one actually finds in the literature do not seem as algorithmic and precise as Jackson's reasoning apparently requires. They explain the sorts of factors that supposedly ground content relations, but they rarely if ever provide formulae which one could use to mechanically compute intentional contents. More often what they offer is a framework within which interpretation can take place, but interpretation is seldom if ever a matter of simple computation or deduction. Jackson might dismiss this as merely a practical shortcoming of our current incomplete theories, a lapse that could be remedied with further refinement. Perhaps so, but one should be cautious about using our current naturalist theories to support belief in any such future ideal theory. It is not obvious that our current models are of a sort that could ever have the requisite precision and computational rigor.

Third and most importantly, the extant naturalistic theories of intentionality may not address the representational facts of most importance to the KA, namely those that concern phenomenal representation. Jackson acknowledges that there is a big difference between representing a given content in thought and phenomenally representing it in visual experience. Thinking the tomato is red is not the same as seeing that it is. Part of the difference is no doubt a difference in total content - the visual experience represents far more than the tomato's color - but that would seem to be only part of the difference. The structure and organization of phenomenal consciousness almost certainly plays an important role in determining the intentional content of conscious states, yet none of the major current naturalistic theories of content has much if anything to say about its role, at least not explicitly. Of course, it may be

possible to expand those theories to accommodate such phenomenal factors, but at present that is more a hope or a promise than a theoretical reality, given their relative silence on the critical factors. Thus confidence in the possibility of constructing such a theory likely depends on more general considerations rather than on the actual specifics of current models.

Jackson's belief in the deducibility thesis relies in part on his general *a priorist* view about physicalism and the realization relation. If entities, facts or properties at some lower level are sufficient to produce those at a higher level, he believes it must be possible at least in principle to *a priori* deduce the latter from the former. He takes the metaphysical link of realization in itself to guarantee the epistemological link of *a priori* deducibility. Jackson acknowledges that many physicalists think otherwise, but he professes to be puzzled about why they regard the claim of *a priori* deducibility as extreme. He takes it as essential to validating the belief that the higher order entities or properties are real and physical. As Jackson sees it, if their nature and existence could not be deduced from the totality of the micro-physical facts, the physicalist ought simply to deny their reality. Indeed in "Mind and illusion" Jackson defines physicalism, or what he calls "bare physicalism", as the thesis that

"the world is exactly as required to make the physical account of it true in each and every detail but nothing more is true of this world in the sense that nothing that fails to follow *a priori* from the physical account is true of it." (2004)

Thus for Jackson the deducibility of the intentional and representational facts from the micro-physical facts is just a special case of the general requirement on the intelligibility of physicalism and the thesis that everything real is physically realized. To put it in a slogan, "No realization without *a priori* deduction" or perhaps "No metaphysical determination without logical derivation."

These are very large issues and certainly not ones that I could hope to settle here. However, it is important to see the major role they play in the Knowledge Argument. Indeed from the perspective of the nonreductive physicalist, the illusion that drives the KA is more a matter of its commitment to *a priorism* rather than its failure to accept the RTS or its belief in

phenomenal properties. Mary's learning something upon release provides an argument against physicalism only if the truth of physicalism entails the *a priori* deducibility of all physically realized facts and propositions from those at the micro-physical level. If one thinks of physicalism in terms of a logical empiricist model of reduction and the unity of the sciences involving bridge laws and theoretical derivability, then the truth of physicalism would enable Mary to infer all the physically realized facts. But many physicalists, especially those of a nonreductive bent, reject that account of what physicalism requires.

The application of these issues to the KA is complicated by the fact that Jackson appeals to *in principle* deducibility, while our intuitions about Mary more likely reflect what she or any human agent could infer *in practice*. Jackson writes,

“when I talk of being able to move *a priori* from the physical account to, say, Carter being a one term President, I do not mean being able to move literally, I mean there exists an *a priori* entailment.” (2004)

Thus he couches the thesis of physicalism in terms of a logical relation, but the Mary argument turns on matters of epistemology. It is what Mary supposedly could not *know* that provides the argument's basis for rejecting physicalism. The issue is further clouded by Jackson's use of “*a priori*” to qualify the nature of the entailment. *A prioricity* is an epistemic feature not a logical one. One can use the phrase “*a priori* entailment”, as Jackson seems to do, to mean “logical entailment knowable *a priori*”. But in doing so, one mixes logical and epistemological factors which one must take care not to blur together. If facts or propositions of one sort logically entail or determine those of another, it does not automatically follow that any actual cognitive agents could carry out the relevant deduction or derivation, nor does Jackson claim otherwise. But if one gives up the automatic move from matters of logic to those of epistemology, it is not clear why Mary's apparent lack of knowledge should concern a physicalist. Our intuitions about the gap in her prior knowledge need carry no negative implication regarding the logical dependence of the mental on the physical.

Thus the central conflict behind the KA may not be that between supporters and

defenders of the RTS, but that between those who take an *a priorist* view of physicalism and those like nonreductive physicalists who do not. The latter disagreement is clearly prior to the former, regardless of one's preferred mode of critical reply. Unless Jackson was already committed to *a priori* deducibility, the RTS would not suffice to support his claim that pre-release Mary lacked no propositional knowledge. The RTS serves only to counter the phenomenal properties objection to that claim. Indeed it is not obvious that the *a priorist* even needs the RTS. Perhaps rather than denying the existence of phenomenal properties, the *a priorist* physicalist might concede their reality but claim that they too are physically realized properties, and thus at least in principle *a priori knowable* on the basis of the underlying micro-physical facts. We may at this point have little intuitive grasp of how such an *a priori* deduction might go, but refuting the KA supposedly requires only that it be *in principle* possible. If one is already strongly committed to *a priorism*, one might argue that deducing propositions about intrinsic phenomenal properties of experiences from their micro-physical bases is no less possible in principle than deducing their representational intentional properties. I do not mean to endorse such a claim, especially since my own sympathies are not particularly *a priorist*, but the mere possibility of making such a move, may further serve to show how relatively little Jackson's critical stance depends on the RTS as opposed to on his *a priorist* perspective.

Physicalists who reject *a priorism* can reply to the KA with even less (or no) reliance on the RTS. Their form of ontological physicalism is compatible with Mary's lacking some propositional knowledge about realized properties despite her comprehensive knowledge of those about their physical realization base. Such a physicalist might part company with the KA at Q0, Q3, Q4 or Q5. Wherever he draws the line, he will do so in part because he rejects the view that physical theory constitutes a universal cognitive tool sufficient even in principle for representing and understanding everything that is real in a physically realized world.

The disagreement between *a priorist* and nonreductive physicalists may reflect an even more basic difference in perspective about the nature of understanding and the cognitive role of theories. If one thinks of theories as descriptions or pictures of reality, then it might seem that

the difference in theoretical representations at different levels is primarily a matter of resolution. Those at the base level that describe individual micro-physical facts might correspond to unit pixels, and every representation at a higher theoretical level would in some sense concern patterns or structures constructed out of those pixels and their relations. Fix the pixels and you've represented all there is to represent; at most one needs to do a bit of deduction to read those further contents off the global base level representation. I do not say any actual *a priorist* would literally make such a claim, but it provides a useful metaphor for thinking about their perspective.

In contrast the nonreductivist, stresses the practical role of theories in enabling us to causally and cognitively engage the world in its many aspects and levels of order and dynamic organization. Theories, models and systems of representation are thought of as “cognitive tools” that support modes of engagement that in their respective ways count as successfully understanding real aspects of the world. There is no presumption that the representational structures that are of adaptive value for engaging a given level of organization L_i should be decomposable into (or constructible out of) those that suffice for dealing with the lower levels of reality L_b that provide the realization base for those at L_i . The sorts of representations and cognitive tools that are apt for a given task of understanding will depend not only on the underlying structure of the objects of understanding but also on that of the understanding subject and the various modes of access and interaction that are afforded by their context of engagement.

Of course, even the nonreductivist owes us some account of how the various levels and forms of order fit together and how those at higher levels might be realized by those at lower ones. But that account will likely stop far short of outright deductions or translations between systems of representation at distinct levels. The nonreductivist does not typically expect to find such tight links and constraints between the multiple systems of representations that he regards as pluralistically needed for the dealing with the world in all its diverse complexity (Van Gulick 1992, 2002).

Moreover, given the existence of cases, such as those involving conscious mind-brains,

in which the understander and the understood are internal to one and the same system, the nonreductivist would expect there to be at least some physically realized subjective facts. As noted above in section three, some physicalists accept the existence of facts or propositions that are physically realized but nonetheless subjective in the Nagel sense that they are fully knowable or understandable only from a limited range of perspectives associated with the ability to have particular forms of phenomenal experience. The existence of such facts falls quite naturally out of the nonreductivist view of understanding and representation as contextually situated activities conditioned by the causal and interactive structure of the cognitive engagement between understander and understood. It is hardly surprising that the higher level forms of engagement in self-understanding systems involve modes of access and interaction, as well as the contents supported and determined by them, that cannot be replicated by any intentionally equivalent representational structures built solely with the resources of micro-physical theory or any other third person system of theoretical representation. That is just what a nonreductivist would expect given his view of theories and models as cognitive tools to be applied in particular contexts of pragmatic engagement.

Given their differing perspectives on the nature of theories and understanding, *a priorist* and nonreductivist physicalists part company with the KA at rather different points. Despite both endorsing physicalism in one version or another, they differ in where they respectively agree and disagree with the proponent of the Knowledge Argument. Both the proponent and the *a priorist* critic agree that physicalism requires the *a priori* deducibility of all true propositions about conscious mentality from the micro-physical facts. And thus they also agree on the following conditional: If physicalism is true then pre-release Mary can know all the facts about seeing red (as well as its transpositive: if she can not know them all then physicalism is false.) They disagree only about the state of Mary's knowledge. The *a priorist* critic, such as current Jackson, argues for Mary's being able to infer her way to complete prior propositional knowledge and uses the RTS to rebut the Phenomenal Properties Objection that aims to show otherwise. The proponent of the KA, such as Jackson pre-1998, accepts the same conditional but

denies its consequent and thus claims to refute its physicalist antecedent. Mary can not know or *a priori* infer all the facts, and thus physicalism is false.

By contrast, the nonreductivist critic rejects the conditional and the deductive thesis on which it relies. The nonreductivist denies that ontological reduction via physical realization carries any epistemic implications about our ability in practice or in principle to deduce every physical fact from a physical theory representation of the base level micro-physical facts. Thus the nonreductivist denies that physicalism is refuted or even weakened by the incompleteness of Mary's prior propositional knowledge or by her increasing or enhancing her propositional knowledge and understanding when she first experiences red.

Indeed as noted above, some nonreductivists (I among them) take the existence of physically realized subjective facts to follow naturally from nonreductive physicalism and the pragmatic view of mind and understanding it embodies. Mary's epistemic gain might be taken to confirm rather than refute any such version of physicalism. The theory itself predicts the existence of limits on what pre-release Mary can know and understand about physically realized experiences. Only after having them herself is she able to engage the relevant aspects of reality in the cognitively required way. Thus a nonreductivist who believes in physically realized subjective facts would be most likely to challenge the KA at Q0 or Q5, since these critical options appeal in their complementary ways to just such facts.

Other nonreductivists might reject the deduction thesis on general grounds, but be less sure about the existence of physically realized subjective facts. They might thus be more inclined to locate their specific disagreement at Q3 or Q4. Nonreductivists of all types agree among themselves in rejecting the deducibility thesis and the epistemic conditional that it implies. Their differences mainly concern the varying respects in which they believe Mary could gain new propositional knowledge without refuting physicalism.

6. Summary

Let us take stock of where matters stand on the relation between the RTS and the KA. As we have seen, there are good reasons to think the RTS is neither necessary nor sufficient for

refuting the KA. If one is a physicalist and *a priorist*, as Jackson currently is, then one might use the RTS to block the Phenomenal Properties Objection. But one must still rely upon a strong *a priorist* assumption about the deducibility of all the representational intentional facts to make one's case that Mary lacks no prior propositional knowledge. The really "heavy lifting" seems to be done more by the *a priorist* view than by the RTS. Indeed, as we noted above, if one were strongly enough committed to *a priorism*, one might be able to forgo the RTS altogether and argue that even if experiences have phenomenal properties they too are deducible in principle from the micro-physical facts.

If one is not an *a priorist* and inclines more toward nonreductive physicalism, the RTS will likely play an even smaller role, if any, in one's critical response to the KA. As we say just above, the nonreductivist has several options. If he believes in the existence of physically realized subjective facts, he can draw the line at Q0 or Q5. He can either reject the initial stipulations of the Mary case as inconsistent at the outset, or hold that Mary's comes to grasp such subjective facts when she first experiences red and acquires the requisite abilities to articulate the structure of logical space and thus to understand her location in it. Alternatively, the nonreductivist could concede that Mary gains new propositional knowledge, but only by coming to know a previously known proposition in a new way or to know a new fine grained proposition. According to the nonreductivist, neither of the two would pose any threat to physicalism, and neither seems to turn in an special way on the truth of the RTS.

My aim in his paper was not to evaluate the plausibility or truth of the RTS, but only Jackson's claim that it is essential to defeating the Knowledge Argument and central to seeing where it goes wrong. From Jackson's perspective the RTS is clearly both, but other physicalists, especially those not committed to his strongly *a priorist* view will likely see things differently. There appear to be a variety of alternative plausible ways in which physicalists might block the Knowledge Argument without needing to accept the RTS. And nonreductive physicalists are more likely to locate the central flaw of the KA in its *a priorist* view of what physicalism requires than in its rejection of the RTS or its belief in phenomenal properties. From the

nonreductivist perspective the crucial illusion does not concern phenomenal properties but *a priori* deducibility. Where one sees illusion depends in part upon the lens through which one looks.

Let me end by trying to pull the diversity of critical options into a few coherent overall positions, which may serve to better locate Jackson's critical stance relative to others. If one is a physicalist of any sort, one must resist the conclusion of the Knowledge Argument, but which means of blocking it seems overall most plausible or attractive will depend largely on how it well fits with the particular version of physicalism one holds. There are different "package deals" from which the physicalist might choose, including at least the following two.

#1 The Jackson Package.. From Jackson's perspective, physicalism carries an unavoidable commitment to *a priori* deducibility. He rejects any view of physical necessitation without *a priori* entailment as failing to adequately explain how the mental could exist as a real part of the physical world and why one should not simply take an eliminativist stance and deny its existence. Jackson discerns no basis on which a physicalist should regard anything as real unless its existence and nature is *a priori* entailed by the physical facts and deducible from them. Thus no refutation of the KA is acceptable unless it allows Mary at least in principle to deduce every true proposition about conscious experience from the micro-physical facts.

Jackson finds the representational theory of sensory experience independently attractive, and he offers it to the physicalist as a means to reinforce the deducibility thesis or at least as a shield to defend it from objections that appeal to phenomenal properties such as PHRED-ness. According to Jackson and other representationalists, sensory experience has fewer mental properties than commonly believed and thus fewer properties that need to be accommodated in the physical scheme or deduced *a priori* from the micro-physical facts. Positive beliefs or intuitions to the contrary about the existence of phenomenal properties are explained away as the result of cognitive illusion.

Moreover, the difference between representing a given content in thought and phenomenally representing it in sensory experience is explained as partly a difference in the total

intentional content of the relevant states, and partly as a matter deploying a different system of representation that gives one some new self-directed know-how but no gain in propositional knowledge about the nature or properties of one's own mental, not even about those involved in the use of the relevant representational system. Jackson adopts the Ability Reply and explains Mary's change in epistemic state as solely a matter of gaining such new know-how; again an intuitions to the contrary are rejected as illusory.

#1 Physicalist Refutation of the Knowledge Argument that includes:

- Truth of Physicalism in its *A Priorist* version.
- Truth of Deducibility Thesis (Everything real deducible from the micro-physical).
- Truth of Representationalism/ Representational Theory of Sensory Experience.
- Non-existence of Phenomenal Properties.
- Phenomenal Intuitions are Illusory.
- Mary has Complete Propositional Knowledge before her release.
- Accept Ability Reply. Mary gains only new know-how.

The Jackson Package

Figure 3

#2 The Nonreductivist Package. Other physicalists see the options differently. Nonreductive physicalists in particular find principled grounds for resisting the deducibility thesis, and they deny that an ontological commitment to physical realization entails an epistemological commitment to the capacity of physical theory to serve as a universal means for knowing and understanding everything real. The nonreductivist regards representation, understanding and knowledge as all having an essentially pragmatic aspect. Thus he views them as contextual matters that are conditioned in part by the causal structure of the cognitive agent, as well as by that of the object of understanding and their engagement. Given the dependence of content on such pragmatic contextual factors, the nonreductivist does not expect the relations between disparate systems of representation to typically be matters of strict equivalence, translatability or deducibility as more logical empiricist or *a priorist* models would require.

The nonreductivist regards know-how and propositional knowledge as less distinct and more interdependent than do supporters of the Ability Reply. The acquisition of new cognitive skills and new modes of interaction and engagement can enhance one's propositional knowledge and give one the ability to represent and understand propositions one could not previously grasp. Indeed according to the nonreductivist that is just what happens when Mary first experiences red and as a self-understanding physical system first reciprocally engages those aspects of her own dynamic organization that realize her having such experiences. Thus according to the nonreductivist, physicalism does not conflict with the belief that Mary's newly acquired skills and modes of engagement give her the ability to increase her propositional knowledge, either by understanding old propositions in new ways or even by understanding new propositions about experience.

Indeed the nonreductivist may be able to concede even that Mary learns new coarse-grained propositions without contradicting his physicalism. If nonreductivist physicalism allows, or indeed entails, the existence of physically realized subjective facts and propositions, then prior to release Mary could not even in principle understand or know such facts about red experience. Thus the nonreductivist might either deny that Mary could in principle know all the

physically realized facts before her release, or he might argue that her learning new propositions about physically realized subjective facts pose no threat to physicalism. They are alternative ways of making the same point.

As to the RTS and the existence of phenomenal properties, the nonreductivist is largely free to decide that issue on its individual merits. He need not adopt the RTS or deny the existence of phenomenal properties in order to block the KA. If he finds the specific arguments for representationalism compelling, he can combine it consistently with the rest of his view. However, if he is reluctant to dismiss our positive intuitions about the existence of phenomenal properties as mere illusions, he can try to find a place for them in his model without undercutting his response to the Knowledge Argument. Thus the overall package might be summarized in Figure 4.

#2 Physicalist Refutation of the Knowledge Argument that includes:

- Truth of Ontological Physicalism interpreted as a matter of Realization. (Everything real is physically realized.)
- Denial of Theoretically/Representationally Reductive Physicalism. (Physical theory is Not a universal cognitive means for understanding and representing everything real.)
- Denial of Deducibility Thesis (Not everything real deducible from the micro-physical).
- Pragmatic/Contextual account of Representation & Understanding.
- Interdependence of Know-how and Propositional Knowledge.
- Mary's New Skills and Modes of Engagement Enhance her Propositional Knowledge.
- Existence of Physically Realized Subjective Facts.
- Intuition that Mary Gains New Propositional Knowledge Preserved as Consistent with Physicalism.
- (Neutral on the RTS and the Existence of Phenomenal Properties.)

The Nonreductive Package

Figure 4

These two packages have their respective strengths, and each will be attractive to some subset of physicalists as a means of refuting the KA. Jackson chooses Package #1 and claims it is the only real option. I prefer #2 and so disagree with his uniqueness claim. Nor is it likely that these two exhaust the range of plausible overall combinations. A popular American TV game show of a few years back offered contestants a choice of prize packages behind one of three doors. The key feature of “Let’s Make a Deal” was that players had to decide whether to trade away their prize already in hand for what might lie behind a yet unopened door. Perhaps a even better package awaits the physicalist behind door #3, but then perhaps he might be wise to stay with that available in Package #1 or #2. I leave it up to you to make the best deal that you can.

References

Chalmers, David (1996) *The Conscious Mind*. New York: Oxford University Press.

Churchland, Paul (1985) "Reduction, qualia, and the direct introspection of brain states." *Journal of Philosophy*, 2-28.

Descartes, Rene (1642) *Meditations on First Philosophy*, Paris.

Hardin, C. L. (1988) *Color for Philosophers*. Indianapolis, IN: Hackett Publishers.

Horgan, Terrence (1984) "Jackson on physical information," *Philosophical Quarterly* 34, 147-83.

Jackson, Frank (1982) "Epiphenomenal qualia," *American Philosophical Quarterly*, 127-36.

Jackson, Frank (1986) "What Mary did not know," *Journal of Philosophy*, 291-5.

Jackson, Frank (1998) "Postscript on qualia," in F. Jackson *Mind, Method and Conditionals*. London: Routledge.

Jackson, Frank (2004) "Mind and Illusion," in P. Ludlow, Y. Nagasawa, and D. Stoljar *There's Something about Mary*, Cambridge, MA: MIT Press.

Lewis, David (1988) "What experience teaches," *Proceedings of the Russellian Society*, Sydney, Australia. Reprinted in Lycan 1990a.

Loar, Brian (1990) "Phenomenal states," in *Philosophical Perspectives Vol. 4: Action Theory*

and Philosophy of Mind. Atascadero, CA: Ridgeview Publishing.

Locke, John (1689) *An Essay on Human Understanding* .

Lycan, William (1990a) ed. *Mind and Cognition: A Reader* . Oxford: Blackwell.

Lycan, William (1990b) “What is the subjectivity of the mental?” in *Philosophical Perspectives Vol. 4: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview Publishing, 109-30.

Nagel, Thomas (1974) “What is it like to be a bat,” *Philosophical Review* 83, 435-

76.

Nemirow, Lawrence (1980) Review of T. Nagel *Mortal Questions*, *Philosophical Review*, 475-

76.

Nemirow, Lawrence (1990) “Physicalism and the cognitive role of acquaintance,” in W. Lycan (ed.) *Mind and Cognition*. Oxford: Blackwell, 490-99.

Tye, Michael (1986) “The subjectivity of experience,” *Mind* , 1-17.

Van Gulick, Robert (1985) “Physicalism and the subjectivity of the mental,” *Philosophical Topics*, 51-70.

Van Gulick, Robert (1992) “Nonreductive materialism and intertheoretical constraint,” in A. Beckermann, H. Flohr and J. Kim (eds.) *Reduction and Emergence*. Berlin: DeGruyter.

Van Gulick, Robert (1993) “Understanding the phenomenal mind: are we all just armadillos?” in M. Davies, G. Humphrey (ed.) *Consciousness*. Oxford: Blackwell.

Van Gulick, Robert (2003) "Maps, gaps and traps", in Q. Smith and A. Jokic (eds.) *New Essays on Consciousness*. Oxford: Oxford University Press.

Van Gulick, Robert (2004) "So Many Ways of Saying 'No' to Mary" in P. Ludlow, Y. Nagasawa, and D. Stoljar (eds.) *There's Something about Mary*. Cambridge, MA: MIT Press.