So Many Ways of Saying No to Mary

Final prepublication version. Published in P. Ludlow, Y. Nagasawa & D. Stoljar (eds.)
There is something about Mary: Essays on Frank Jackson’s Knowledge Argument,

Robert Van Gulick
Syracuse University
RNVANGUL@syr.edu
So Many Ways of Saying No to Mary

I. Introduction.

The Knowledge Argument (Jackson 1982) is deceptively simple but radical in aim. We are asked to imagine a hypothetical history of Mary the super color scientist who is held and then released from achromatic isolation. Our intuitions about what Mary would or would not know before and after her release are then alleged to show that physicalism is false, i.e. to refute it by purely a priori means rather than empirically with actual factual evidence.

One is rightly cautious in accepting so radical a conclusion based on merely hypothetical reasoning. Whether or not it is ever possible to achieve so sweeping a result by such austere means, it is surely seldom so and the standard of proof is justifiably set very high: both for the argument and any assumptions on which it relies. One must always weigh the competing plausibility or implausibility of the argument’s alleged radical conclusion against that of its assumptions. If there is a logical conflict of intuitions with established theories, the more empirically well supported the theories or models, the more likely we are to assign the blame to our intuitions and defer to evidence and theory.

Nonetheless the Knowledge Argument has a strong intuitive appeal. It is easily understood even by introductory students, and it provides a tangible vehicle for articulating a deep seated unease about physicalism that may or may not indicate an a priori defect in the position. It has generated philosophic controversy since it was first proposed, and it has continued to do so, surviving even renunciation and disavowal by its original author, Frank Jackson (1998, this volume).

Thus it seems apt to take yet another look at the status of the Knowledge Argument and the assumptions and intuitions on which it and its critics respectively rely.
If it is faulty as so many think it is or must be, one would like to have a good explanation of just where it goes wrong and why it still holds such appeal if it is in error. Either it is right, in which case it surely deserves our attention, or it is wrong and we should then like to better understand how it generates such a powerful illusion.

Ten years ago, (Van Gulick 1993) I surveyed the state of play at that time and judged its critics to have gained the upper hand by finding multiple weak points in its argumentative armor. I offered a set of diagnostic questions to sort the various ways in which one might part company with the argument’s reasoning, and found several such departure points sufficiently compelling to give one good grounds for rationally resisting its conclusion. That assessment was not universally shared, and even those who might agree in ultimately rejecting the argument and its conclusion might still find the replies that I surveyed not adequate to disarm it nor pinpoint its key defect.

Thus I will begin by briefly reviewing the main lines of response I earlier identified. I believe the questions used ten years ago still distinguish relatively well among the main critical variants, though with a few qualifications that will be noted afterward. Two sorts of revision of the map are needed. On one hand, there seem to be additional ways of disputing the argument and perhaps of “getting off the train” even earlier. On the other hand, by looking further into the details of some of the prior responses we can see some links between them that might not have been obvious before, and as a result some of the original distinctions may begin to blur a bit. Thus I will follow the quick review with the relevant additions and extensions.

As noted above, Jackson himself has recently come to reject the Knowledge Argument (1998). In preparing the final draft of this paper, I had the opportunity to read Jackson’s explanation of his reasons for doing so (this volume), and thus I can now also locate his own current view within the space of critical options, which I will do in my final section.
II. A quick digression on boomerangs.

Before reprising that taxonomy of critical replies, I should note that the Knowledge Argument is a species of what I have elsewhere (Van Gulick 2003) called “boomerang arguments.” (The label is chosen partly in jest of the fact that such arguments have been prominently offered in different forms by Jackson and by David Chalmers, both Australians). The distinctive feature of a boomerang argument is that it reaches across to the epistemic/cognitive/conceptual domain of facts about our representation of the world, and then swings back to reach a conclusion in the metaphysical/ontological/factual domain about the nature of reality itself. It moves from facts about how we represent or conceptualize the world to supposed results about the necessary nature of the world itself. More specifically, boomerangs often move from supposed gaps or lacks of links 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. (Van Gulick 2003)

The mode of argument has a long but dubious history: there are so many ways in which we might think, represent or conceptualize two things as distinct and yet be wrong because they really are the same in ways not apparent to us. The fault or difference often lies not in the world but in our concepts of it. Nonetheless, boomerang arguments have been popular for a long time.

Right at the seventeenth century start of what we call “modern philosophy” Descartes’ (1642) attempted in Meditation VI to prove the nonidentity of mind and body by appeal to the modal difference in one’s supposed ability to imagine one’s continued existence in the absence of one’s body but not in the absence of one’s mind. Of course, as we well know, many things that seem possible or impossible at one stage of scientific knowledge turn out just the opposite with further knowledge of the real structure of the world.
We can not have water without having $\text{H}_2\text{O}$, nor an increase in the heat of a gas without an increase the kinetic energy of its molecules since, as we have known since the rise of modern chemistry and the demise of caloric theory, that it is all that water and the heat of a gas are. The heat of a gas is not some correlate or consequence of its kinetic energy but one and the very same thing. So any attempt to imagine the one without the other is bound to result in a contradiction whether apparent to the earnest imaginer or not. What we imagine at most is a world in which something that is not heat nonetheless shares some of heat’s superficial properties. They might be some of those that go to make up what Locke (1689) called its “nominal” as opposed to its “real” essence, or what in more contemporary two dimensional semantics might be identified with the primary intension of the term “heat (of a gas)”. Perhaps we imagine a world in which something other than molecular kinetic energy affects our thermo-receptors in a way that mimics heat, as some believe the capsaicin of “hot” peppers does in this world. That we can do. What we can not do is imagine a world in which a gas has an increase of heat (as the kind is picked out by the secondary or contextually determined intension of “heat”) but has no increase of molecular kinetic energy.

Prior theories and concepts may not have ruled out their being independent and distinct, but actual empirical discovery reveals that there is but one thing in the world to which those two concepts both apply in their partial though effective ways. Thus we now know that real heat and real kinetic energy - the real things or properties in the world at which our concepts aim and to which we refer in word and thought - are not capable of independent existence because they are just one thing.

Similarly Descartes’ boomerang begs the question against the physicalist. The separate existence of mind without body is conceivable in the strong sense needed to license a claim of logical possibility and the move to actual distinctness only if mind is in fact distinct from body. If the brain and mind should be one in the same, then
Descartes’ attempt to imagine the one remaining in the absence of the other must lead to a contradiction. Such a separation would not be logically possible, regardless of how possible it might have seemed from his limited conceptual and epistemic perspective. If the lack of entailments between our concepts were sufficient to validly infer the independence of the items to which they apply, doing science would be a lot easier. But of course we know it is not. What looked to Descartes to be an obvious and a priori modal difference need not in fact reflect any real difference in the world, but merely the limits of his understanding.

These criticisms are well known, but they have not in themselves sufficed to prevent new boomerangs from being thrown. Their current wielders are confident that unlike Descartes’ mistaken argument their own attempts to reach results about reality from mere conceptual and epistemic intuitions can deflect the sorts of problems that set their predecessors off course, with various maneuvers including those that invoke two-dimensional semantics. (Chalmers 1996, Jackson 1982) We should remain open to that possibility, but a fair measure of skepticism about any such attempt would seem to be in order.

The knowledge argument is especially vulnerable and under threat of doubly faulting in this regard. Not only does it rely upon a priori conceptual intuitions to make its case, but the relevant intuitions themselves concern epistemic concepts. It is our supposed intuitions about what would or could (or not) be known in various highly counterfactual situations that lie at the core of the argument. Thus the knowledge argument attempts not only to infer results about the necessary structure of reality from purely hypothetical reasoning about highly nonactual conditions, it aims to do so by engaging in such reasoning about supposed epistemic facts of what we would or could not know in various situations. Thus we need first to rely upon a priori intuitions to reach our supposed epistemic facts and then to infer ontological conclusions from those
alleged epistemic limits. If I may push the metaphor a bit, the boomerang has to make a double loop and circle twice within the conceptual epistemic domain before landing back upon the real world objective side. Again that does not mean the trick can not be done, but caution and doubt seem justified, as does the high level of critical scrutiny the argument appears to merit.

III. Resurveying the range of replies

The basic structure of the knowledge argument provides the anchor points for the various critical replies I earlier surveyed (1993). The argument, which relies upon two main premises and two inferences, might be given in simple form as AK~P:

Argument AK~P.

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.

We can sort most of the critical replies to the argument in terms of the answers they give to a series of successive questions that locate the point at which various critics part company with the original line of argument as in figure 1.
Q2. What sort of knowledge does Mary gain? Is it strictly know-how, or does it include new knowledge of facts or propositions.

\[
\begin{array}{c|c}
\text{NO (Churchland, Hardin)} & \text{YES} \\
\hline
\text{X Physicalism not refuted.} & \text{II} \\
\text{II} & \text{V}
\end{array}
\]

Gains only new know-how. Gains new knowledge of propositions.

\[
\begin{array}{c|c}
\text{Lewis, Nemirow} & \text{II} \\
\hline
\text{X Physicalism not refuted.} & \text{II} \\
\text{V}
\end{array}
\]

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

\[
\begin{array}{c|c}
\text{NO. She only comes to know old propositions in a new way.} & \text{YES. She comes to know new propositions.} \\
\hline
\text{Tye, Horgan, Churchland} & \text{II} \\
\text{X Physicalism not refuted.} & \text{II} \\
\text{V}
\end{array}
\]

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

\[
\begin{array}{c|c}
\text{Only on a fine grained mode.} & \text{On a coarse grained mode.} \\
\hline
\text{Lycan, Loar} & \text{Jackson pre-1998} \\
\text{X Physicalism not refuted.} & \text{Physicalism refuted.}
\end{array}
\]

(Each “X” marks a point at which the knowledge argument is blocked.)

Map of Critical Options

Figure 1
IF  •Mary were to know everything physical there is to know about seeing red
  •while never having herself experienced red, and
  •she were then released and allowed to see red,
THEN  What would she learn or come to know?

Different replies to the argument disagree at various points with its story about what comes after the “THEN”. According to the knowledge argument, Mary gains new knowledge of a phenomenal sort, whose existence suffices to refute physicalism. Scope is important here. It is not that Mary comes to know something that suffices to refute physicalism; she has no access to any such proposition, nor does the argument claim she does. What supposedly refutes physicalism is not what she learned, but rather the mere fact that she learned it. For if physicalism were true, it would seem there could be nothing left for her to learn, given the assumption that she already knew everything physical there was to know. The argument thus might be read as an intended reductio of any physicalist claim to exhaust (factual or knowable) reality as in PEx.

PEx. The physical facts are all the facts there are.
The argument contends that were that so, there would not have been anything left for Mary to learn upon her release, but since she does PEx must be false.

Thus the first obvious point for disagreement is in answer to question Q1.

Q1. Does Mary learn anything or gain any knowledge when she first experiences red?
The knowledge argument assumes a positive answer, but some critics dispute that. (Churchland 1985, Hardin 1988)

The argument appeals to intuitions that are supposed to lead us to imagine a scene in which Mary would express surprise and perhaps say something such as, “At last I really know what it is like to see red.” Or if Mary were feeling philosophical she
might add, “Now (and only now) do I know the full subjective nature of red experience.” That at least is how the argument imagines our intuitions going.

Some critics think otherwise. They fault the argument for underdescribing the hypothetical situation in ways that produce off-target intuitions. In particular there is the danger that one’s intuitions will conflate the stipulated conditions with more normal cases of real knowers, whose physical knowledge is vastly more modest than that Mary is supposed to have. Indeed we are asked to assume that she knows *everything* physical there is to know about having a red experience. Such absolutely comprehensive physical knowledge - i.e. complete knowledge of every physical science detail of the relevant brain states including every aspect whether microphysical, chemical, physiological, neuroanatomical or neural-network-dynamical as well as every link and detail in between - is so unlike that which any of us might ever come close to having, that our intuitions in thinking about such an extreme case are easily diverted and lead astray.

The cognitive constraints that operate upon our epistemic concepts and structure our intuitions about how they might apply in hypothetical situations are conditioned heavily by their operation in their actual domain of application and engagement. We can of course extend our reasoning and thought into counterfactual situations, and try to apply our concepts in nonactual domains. But the farther those contexts diverge from our own, the less likely will those concepts apply unproblematically in the full range of encountered cases. Thus caution is in general prudent when reaching intuitive a priori judgments about highly nonactual hypotheticals.

The general worry clearly applies to the specific case of Mary and especially to the unlimited range of physical knowledge we are asked by assumption to impute to her. As a matter of actual human cognitive limits, it may well be impossible for any real human to know all that Mary is supposed to know. It might simply outrun our human
memory capacity, the limits on what we can hold in attention, or our computational and integrative powers; indeed Jackson (this volume) concedes that it would do so. But Jackson and defenders of the knowledge argument regard these as mere “practical limits”. They believe what matters is what can be known “in principle.” What humans cannot in practice know might still in principle be knowable by our cognitive superiors: super cognitive humans freed from our contingent limitations. And yet even SUPER Mary would not know all there is to know. According to the argument’s defenders, Mary’s pre-release limits do not derive from any quantitative limits on how much she can know but on the fact that everything she knows is physical. And so SUPER Mary’s ability to know such things without quantitative limit does not give her any real advantage over Mary in the key respect. More physical knowledge - no matter how extensive or exhaustive - will not do the job. Or so at least that is how the argument’s defenders imagine our intuitions would play.

Some critics feel otherwise. (Churchland 1985, Hardin 1988) They believe that if one describes the hypothetical case fully and emphasizes the highly nonactual and absolutely comprehensive nature of Mary’s physical knowledge, then our intuitions would have to be recalibrated with the result that for many people the mostly likely scenario might become Mary’s lack of surprise and her saying something much more like, “Yes it is just as I knew it would be.” Or waxing more philosophical she might declaim, “This experience fully confirms my predictions and prior inferred knowledge of the phenomenal aspect of seeing red as detailed in my physico-neural model of color experience.” The fact that in real world situations any actual and partial physical stories we might get would leave us in the dark about what it is like to be in such a state should not be projected onto the case of Mary or SUPER Mary without good argument for treating the cases as equivalent.
The hypothetical epistemic situation is so unlike the real world contexts in which we operate that it difficult to have uncontroversial intuitions about it. Can the knowledge argument’s defenders make the case that their imagined scenario of Mary’s post-release surprise is the only a priori acceptable one? Can they assume that the alternative imagined by its critics - that in which Mary proclaims confirmation of her prior expectations - can be ruled out as impossible by pure a priori reflection on our concepts? It is a big assumption to make, and one can already feel the counterbalancing of plausibility between the argument’s assumptions and its radical conclusion. The argument’s need for a decisive intuitive judgment in favor of its preferred scenario may seem a thin reed on which to hang so momentous a metaphysical result.

Thus some critics may feel justified in parting company with the argument already at this preliminary stage by simply answering “No” to question Q1. They may believe that given a proper appreciation of the extent and comprehensiveness of Mary’s prior physical knowledge the proper reply is, “No, Mary would not learn anything or come to any new knowledge. She would have known it all already.”

A negative reply to Q1 would stop the knowledge argument at the outset; thus it offers a first potential point for parting from its line of reasoning. Since my aim for now is to review my original survey, I will follow it by leaving matters with Q1 thus unresolved and open to conflicting intuitions. However, when we turn below to amending that review, we will find that there may be ways of departing from the argument even earlier by challenging P1. Moreover, we will also need to consider ways in which supporters of the argument - though not Jackson himself (1998) - might defend their positive reply to Q1 and their favored intuitions by appeal to a seemingly plausible empiricist thesis about the necessarily experiential origin of some sorts of knowledge, which we can call
in deference to John Locke (1689) “The Pineapple Principle.” However, we will put off those matters till later in Section IV after we have completed our survey in review.

Most critics of the knowledge argument are willing to concede that Mary learns something upon her release but still deny that its conclusion follows. Their disagreements arise further along in the argument’s reasoning, which thus brings us to our second diagnostic question Q2.

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

One prominent line of reply to the knowledge argument, originally proposed by Lawrence Nemirow (1980, 1990) and supported by the late David Lewis (1988) concedes that Mary gains knowledge upon her release but insists that all she gains is added know-how. According to this so called “ability reply” Mary gains new skills and abilities to imagine and recognize certain states as red experiences, but she gains no new propositional knowledge. If we follow Lewis in regarding a gain of propositional knowledge as a reduction of uncertainty and a narrowing of one’s knowledge of one’s location in the space of possibilities, then according to the ability reply Mary gains no such knowledge upon her release. Her exhaustive physical knowledge already suffices to eliminate any uncertainty about the nature of her state or her location in possibility space. If physicalism is true, then once the physical facts are set there are no genuine options or possibilities left open; one’s location has been fully determined.

Mary’s state is surely different after her release. Her brain has actually now been in the physical state that is identical with seeing red, and as a result of having been in it she may well acquire practical abilities to reproduce that state again in herself through an act of purposeful imagination or to recognize it as the same (or nearly so) should it (or some similar state) reoccur. These skills may well be ones that Mary lacked before. Her mastery of the physical descriptions of such states - no matter how detailed and
complete - need not have given her any such abilities to produce or recognize their occurrence in her own brain from the inside. Knowing the physical details of some gastric state does not by itself enable me to produce it in my stomach.

Brains, for obvious reasons, have significant abilities to put themselves back into prior states and to distinguish among various states that they have been in, but many such abilities seem to depend upon having been in the relevant state at some prior times. The system’s ability to put itself into the state is a matter of its being able to get back to some state that has left a causal imprint on its structure and dynamics. Perhaps the brain’s network activity has something to do with it. Networks, like those in the brain that subserve learning, often show a tendency to move back into prior activation patterns. The networks show “biases’ in favor of some such past patterns of activity.

Whatever the underlying explanation, the basic facts of memory, imagination and recognition are obvious from the first person point of view and accepted as matters of folk psychology. Having seen or heard something novel, one typically gains some ability to call it back through an act of memory or imagination and to recognize like states when they arise on subsequent occasions. So it seems uncontroversial to suppose that Mary upon first seeing red gains some such practical abilities and skills. Thus at least part of what she gains is practical know-how, but the key question is whether or not that is all that she gains. Advocates of the ability reply say, “Yes, that is all she gains,” while others both among the argument’s supporters and its critics disagree and claim, “No, she gains more than just new know-how.”

Once again we are left with a clash of epistemic intuitions about our highly hypothetical case. Supporters of the Ability Reply may defend their preferred intuitions by portraying the contrary view as mistaking the mere acquisition of a new mental skill as a gain in factual or propositional knowledge. They will try to get others to see that what looks like new propositional knowledge is just an illusion or confusion about Mary’s
new abilities. However, they seem unlikely to persuade many of those who disagree to switch their intuitions. The pull of the contrary view, according to which Mary comes to some new propositional or factual knowledge about the nature of red experience and what it is like, seems too compelling to dispel by mere appeal to alleged illusions that supposedly confuse new abilities with gains in propositional knowledge. Nor are attempts at swaying intuitions in the contrary direction any more likely to succeed.

Thus we seem left at an impasse of conflicting intuitions with little present hope of decisive resolution. Finding a way round the stalemate will likely require some rethinking of the issue that places the specific question asked by Q2 in a larger context, one that offers a different perspective on the issue and thus perhaps provides some insights that enable us to see our way more clearly through the thicket of conflicting intuitions. I will consider some such possibilities below in Section IV when we get around to extending the original survey, but let me first continue my review by summing the state of play with Q2 and then moving on to our next diagnostic question.

If we answer Q2 as the supporters of the ability reply do, then the knowledge argument is blocked and its anti-physicalist conclusion does not follow. If all Mary gains is new know-how, then her prior comprehensive physical knowledge need not have left her with any lacks in propositional or factual knowledge, and the physicalist account of what is real has not proven incomplete. However, the question of whether that is all she gains remains open to conflicting intuitions. If one answers Q2 to the contrary and holds that Mary gains more than just new know how, then one will need to move farther into the argument: either all the way to its conclusion or to some further point down the line at which one finds some other reason to depart, which brings us to Q3.

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

Q3 might seem to be redundant and to be simply re-asking that already posed by Q2, but there is an important difference indicated by the shift in what “new” modifies. Q2
asks whether Mary gained any new knowledge of a propositional or factual sort, whereas Q3 inquires as to whether Mary has gained knowledge of new propositions or new facts. In Q3 the “new” applies to the facts and propositions known, whereas Q2 concerns “new knowledge.”

The difference might be significant if one believes that there may be enough of a change in Mary’s epistemic state to count it as a new state of knowledge, even if it has the same propositional object as some epistemic state she had already been in before her release. This might be so, for example, if after her release she came to know a previously known proposition in some quite different way. The newness of the knowledge state might turn on some aspect other than the novelty of its propositional object, perhaps on its affording a new mode of access to it. For example, Mary might come to know directly by introspection or acquaintance some fact or proposition she had previously known only indirectly through theoretical inference. Or perhaps Mary gains the ability to represent the relevant proposition in a new system or medium of representation; she might now be able to use some basic biological system of representation to grasp a proposition which she had previously been able to comprehend only through the medium of scientific theory. Yet if the properties and relations that constitute that fact or proposition are the same as those known before, one might justifiably describe her not as learning a new proposition but only as coming to know an old (i.e. previously known) proposition in a new way.

Indeed this is just the line taken by some critics (Horgan 1984, Churchland 1985, Tye 1986) who concede in reply to Q2 that Mary gains new propositional knowledge but deny in response to Q3 that she comes to know any new propositions. That conjunction of replies need not entail a contradiction as long as one is willing to individuate two states of propositional knowledge as distinct on some basis other than their having different propositional objects. If two states of propositional knowledge can share the
same propositional object yet count as different, then there is the possibility of conceding that Mary learns more than know how, but nonetheless does not increase the range of propositions that she knows. One could consistently give a “Yes” to Q2 and a “No” to Q3.

A negative answer to Q3 thus provides an alternative and less obvious way of resisting the knowledge argument. If Mary does not come to know any new propositions, then none need have been left out by the physical story that she knew before release. If the only sense in which she gains new propositional knowledge is that she acquires new ways to access or represent propositions she already knew, then her epistemic change need pose no threat to physicalism. However, it is the innocuous nature of such a change that leads others to reject it as not doing justice to the intuitions that motivate the knowledge argument. David Lewis (1988), for example, argued that it would trivialize the notion of gaining new propositional knowledge. Lewis argued on that such an interpretation would lead to the absurd conclusion that one would gain a wealth of new propositional knowledge simply by learning a new language, whether French or Urdu - that allowed one to represent to oneself in a new way facts one already new. That I can now represent the proposition that snow is white to myself by use of the French sentence, “La neige est blanche,” or the Urdu sentence, “Barf safed hoti hai,” would not seem to add to my stock of propositional knowledge, at least not to that about matters of color or winter precipitation as opposed to any linguistic facts I might have learned about French or Urdu themselves. Such critics dismiss the strategy of drawing the line at Q3 and argue that nothing should count as a gain in propositional knowledge unless it involves coming to know new propositions and thus adding to the range of propositions that one knows. It would be at best misleading to count anything less as “new propositional knowledge,” or so at least they claim.
Once again there is room for disagreement, with proponents of the Q3 strategy on one side and on the other a mix of those who support the knowledge argument and those like Lewis who wish to reject it but on other grounds. Clearly the supporters of the argument intend for Mary to gain new propositional knowledge in the stronger sense of coming to know new propositions. Nothing less would suffice to secure their anti-physicalist conclusion. Nor is it plausible to read the notion so weakly as to count mere mastery of a foreign language as sufficient for gaining new propositional knowledge of previously known matters, whether about the weather or anything else. However, those who aim to draw their line against the knowledge argument with a positive answer to Q2 and negative to Q3 do not seem to be making so simple and obvious an error. Moving from an ability to represent facts about the nature of color experience in the language of theoretical neuroscience to being able to do so through the use of an innate biologically based system of coding seems a far larger and less trivial change than switching from English to French or Urdu in describing facts about snow. Moreover, the change in the system of representation is so much more radical in the former case, that it seems far more likely that one might confusedly regard a new representation of an old fact as a case of actually coming to know a new fact or proposition. Of course those who support the knowledge argument do not intend it in that way, but the aim of its critics is to undermine the argument by showing how the intuition on which it relies might result from error or confusion rather than epistemic reality.

The situation is further complicated by the diversity of ways in which one might individuate propositions. Q3 asks whether Mary comes to know any new propositions, but there is controversy as well as about what should count as distinct propositions. The issues raised by Q3 thus run into those in Q4, the final diagnostic question in our reprised survey:

Q4. On what mode of individuating propositions does Mary learn a new proposition?
Q4 comes into play only if one has given a positive answer to all the prior questions including Q3. If on the contrary, one denies that Mary learns any new propositions Q4 is moot. Q4 provides a seemingly final site in our review at which to disagree with the knowledge argument, one that allows the critic to concede the Mary gains knowledge some of which involves learning new propositions, and yet still resist its final anti-physicalist conclusion. Making such a last minute swerve depends upon individuating propositions in a finely intensional way.

Propositions and facts can be individuated in various coarse or fine grained ways depending upon what range of factors one counts as making propositions the same or distinct. On the coarse side, one might identify propositions with functions from possible worlds to truth values, or equivalently as the range of worlds in which the function for a given proposition assigns the value True. The proposition that snow is white is defined in terms of the conditions under which it is true; any other proposition true under exactly the same and only the same conditions would count as the same proposition. Thus logically equivalent propositions get collapsed into a single proposition. On such a coarse grained mode of individuation, the proposition that 5+7=12 is the same as the proposition that \(\_63 = 3 \times \_7\). They are both necessary mathematical truths and thus both true in every possible world. Since they have the same function from worlds to truth value, namely the necessary function that assigns the value True to every possible world, they count as the same coarse grained proposition. Lewis, who often defined propositions as such functions, held as we saw above that an increase in propositional knowledge requires a reduction in uncertainty and a correlated narrowing of one’s location in the space of possibilities. In that respect as well the two mathematical propositions are the same. Neither locates me within any subdomain of the space of possible worlds, since both hold in all worlds. Of more direct relevance to the mind/brain case is the fact that the coarse grained scheme of individuation also counts
the propositions, “This cup contains water” and “This cup contains H\textsubscript{2}O” as the same.

Given the scientifically established identity of water with H\textsubscript{2}O, whatever is true of water at a world is true there of H\textsubscript{2}O as well since they are one the same substance. Or if one prefers to talk of properties, one can identify the property of “being water” with that of “being H\textsubscript{2}O”; they are one and the same real property whatever differences there may be in the two concepts by which we pick it out. Properties as well can thus be defined as functions from worlds, but here they are taken as functions from world to sets of objects, those that at that world possess that property. On that scheme as well, water and H\textsubscript{2}O everywhere coincide.

Alternatively one might individuate propositions more finely by distinguishing among them in terms of their internal intensional structure, most especially in terms of the incorporated concepts by which they respectively determine their coarse grained functions. 5+7=12 might coincide at its world-to-truth-value function with _63 = 3 x _7, but still be counted as a different fine-grained proposition on the basis of the difference in the constituent concepts by which they respectively construct the overall world-to-value function that they share.

What then of the proposition that the cup contains water and the proposition that the cup contains H\textsubscript{2}O? Are they the same proposition or not on such a mode of individuation? The overall propositions share the same world-to-truth-value function, and the reference function for “water” coincides at every world with that for “H\textsubscript{2}O”. They pick out the same property and the same objects, stuff or regions in every world. They can not diverge since as we said above water just is H\textsubscript{2}O; there is only one real property in the world apprehended from two conceptual modes of access. The issue thus is whether or not mere differences in such access or mode of presentation at the conceptual level should lead us to distinguish among two such propositions. Should the differences between our concepts of “water” and “H\textsubscript{2}O” matter even though they pick out
the same real property? Should two propositions that coincide everywhere except that one employs the concept “water” and the other “H$_2$O” be counted as different or the same? The question has no absolute answer. Both modes of individuation are available and well behaved, the issue is which system of categorization and individuation better fits the context at hand. The sorts of situations that we want to model or explain will to a large part determine which mode of individuation is most apt.

How then should propositions be individuated in the context of the knowledge argument? At a minimum the argument must treat the notion consistently in its various steps to be valid. The claim that Mary learns a new proposition must rely upon the same mode of individuation as does the claim that if the complete physicalist story leaves out any true propositions then physicalism is false. The two claims must agree in how they count propositions as the same or different. If not, the argument’s premises would not connect to support its conclusion.

A coarse grained mode of individuation, e. g. by world-to-truth-value function, would seem to underwrite the latter hypothetical claim about physicalism. Its claim to be a comprehensive account of reality would seem in doubt if it were shown to leave out propositions of the coarse grained sort (though on this point as well questions will be raised in Section IV.) But what if all the physical story left out were propositions that differed in only the fine grained sense from ones that were included? Would physicalism’s claim to completeness be undermined by the mere fact that it does not capture or include every variant conceptualization of the real states of affairs and real properties that it nonetheless succeeds in picking out through purely physical concepts?

It would seem not, at least if physicalism is interpreted along the lines of the sort of nonreductive physicalism that has become in one form or another the mainline position in the contemporary philosophy of mind. Nonreductive physicalism combines (or at least aims to combine) a commitment to ontological physicalism with a pluralist
attitude about theories, concepts and modes of representation. Everything that is real is physical in the sense that it is in some way or another realized by underlying physical processes or structures. But that universality of physical realization is compatible with our epistemic and theoretical need for a diversity of means of representing and modeling reality, including many that fall outside the domain of physical theory and the conceptual and representational resources it provides. Every economic transaction may be fully realized by physical events and processes but no one would ever suppose that the physical sciences provide adequate means for describing, modeling or explaining economic regularities. To think otherwise would be to adopt a form of extreme cognitive or conceptual imperialism about the reach of physical theory. One surely need not buy into such an implausibly strong view about physical theory as a universal cognitive tool in order to count as a physicalist in good standing. Indeed nonreductive physicalist reject that very demand in their commitment to theoretical and conceptual pluralism. So if the only sense in which Mary learns a new proposition is in the fine-grained sense that reflects distinctions of internal conceptual structure, then her acquisition of new knowledge upon release need not conflict with physicalism, at least not if physicalism is interpreted as a basic ontological claim with no pretension that physical science by itself provides all the concepts and modes of representing that we might validly use or require for our diversity of cognitive engagements with the world (Loar 1990, Lycan 1990b, Van Gulick 1985 and 1992) Still as one of the editors of this volume (D. Stoljar) has advised me, a defender of the knowledge argument might object that Mary prior to release should at least in principle be able to infer every fine grained proposition true of seeing red from the totality of strictly physical facts. If we let “P” stand for the physical facts and “E” for the experiential facts, does physicalism entail that the conditional P \rightarrow E is a priori? If it does, then it would seem that Mary should be able to infer the truth of E and thus of any true fine grained proposition about red
experience before her release? Her inability to do so would thus seem to refute in any version of physicalism that entails the a prioricity of P→E. I think the right response is to deny that P→E is a priori, at least if E is taken to include every fine-grained proposition true of experiencing red. The physicalist is committed to the necessity of the conditional but not to its being a priori knowable or to Mary’s being able to deduce E from P. In particular, there may be good reasons, as I will discuss below in section IV, for believing that Mary prior to her release could not command all of the concepts needed to grasp some of the fine-grained propositions included in E. If she can not even grasp the relevant fine-grained propositions, she surely is in no position to infer their truth. Though her conceptual limitations might thus deny her cognitive access to some fine-grained proposition E*, she could know some proposition P* that counts as the same as E* on a more coarse-grained mode of individuation. As long as she knows the truth of P* prior to release, her coming to know E* afterward would not involve any reduction in her uncertainty of a sort that would pose a problem for physicalism. Her inability to infer every fine-grained proposition necessitated by the physical facts need pose no problem for the physicalist. Thus the Knowledge Argument can deduce the falsity of physicalism from the incompleteness of Mary’s propositional knowledge only if incompleteness entails her failing to know some proposition in the coarse-grained sense.

Consistency thus leads us to ask how step A2 of the argument fares if read as relying on that same coarse-grained mode of individuation? Can Mary be said to learn a new proposition in that sense? Does she learn a proposition in the Lewis sense of reducing her uncertainty and fixing herself more locally in the space of possibilities. Or is it only in the more fine-grained sense that she comes to know something new that was earlier left out? Does she gain only a new line of conceptual access to the same states of affairs she already knew? Those who wish to use Q4 to cut the knowledge off late in
the game, clearly believe that the propositions that Mary learns are new only in the fine grained sense. However, supporters of the argument think otherwise and deny that what Mary comes to know about experience of seeing red involves not merely reconceptualization of old facts, but facts not previously known under any conceptualization. Given the highly abstract nature of the matters about which the two sides present their respective intuitions, it is difficult to see what might resolve the conflict. One side sees mere reconceptualization of a coarse fact involving one and the same property apprehended under differing modes of presentation. The other side sees a new fact that picks out a real property not previously included in the physicalist story. So which it is to be: new property and new coarse proposition, or old property grasped through new concept and new fine grained proposition? How is one to decide? Is mere reflection upon what one’s introspective experience would be like in such a hypothetical situation supposed to provide all one needs to decide the conflict? That surely seems unlikely, and this may be yet another point at which one may wish to weigh the competing plausibility of the argument’s assumptions against that of its radical conclusion.

How should someone who is otherwise disposed toward physicalism on scientific grounds weigh the cost of crediting the argument’s intuition that Mary learns a new proposition in the coarse grained sense against the cost of having to reject physicalism? How much greater weight must we assign to the coarse grained intuition over the fine grained one in order for it to outweigh the prior positive value assigned to physicalism? When we have conflicting plausible intuitions, one of whose acceptance would require us to make a major modification in our accepted and otherwise empirically supported theory of reality, it is not irrational to demand that the destructive intuition achieve a clear victory over its competitor in order to win our acceptance of its high cost. The knowledge argument seems unlikely to win any such decisive victory in the conflict of
intuitions raised by Q4. Thus even those, like nonreductive physicalists, who might follow its line of argument through Q1, Q2 and Q3 will be inclined to disagree at Q4. They will claim that Mary learns new propositions in only the fine grained sense, and that no a priori reasoning about our concepts can rule out the possibility that the property of experiencing red that Mary conceptualizes introspectively after her release is just the same as some physico-neural property that she had already earlier grasped through physical concepts. Absent some compelling reason to reject that view in favor of the conflicting intuition, the physicalist can take his stand at the Q4 choice point and rationally decline to take the argument’s final step. Q4 thus provides yet one more basis for declining to accept the argument or its conclusion.

Let me sum up my reprise of the critical survey of options. We asked four diagnostic questions, each of which offered the chance of answering in a way that would suffice to stop the argument and block its conclusion. In each case the answers remain open to plausible dispute and contention. Something can always be said on behalf of the answer required by the argument, but plausible reasons can also always be made for the alternative intuition. Going against the argument’s preferred option in any of the four cases would by itself suffice to disable the argument. One might deny that Mary would learn anything given her vast physical knowledge (Q1). Secondly one might hold that all Mary gains is new know how and new abilities (Q2). A third option would be to claim that Mary comes not to learn any new propositions but only to know old propositions in a new way (Q3). Finally one might hold that Mary learns new propositions but only in the fine grained sense (Q4).

To repeat my invitation of a decade ago, “Choose your favorite. Take your pick.” Those who find physicalism otherwise plausible have ample options for rationally resisting the knowledge argument and its radical contrary conclusion. Indeed given that the argument’s success requires winning all four contests between conflicting plausible
intuitions, it is difficult to understand why physicalists have devoted as much attention to
the argument as they have. It would not seem to pose much of a threat. Perhaps like a
porch light pulling moths off course, it is just too attractive to pass by - even if one is
drawn in only to refute it. In fairness, I should note that two editors of this volume (D.
Stoljar and Y. Nagasawa) have argued to the contrary that the very disunity of
physicalist responses to the Knowledge Argument suggests that physicalists lack a
clear answer to the challenge that it poses. Thus it may be useful at some point in the
next section to consider whether the four seemingly incompatible responses I have
surveyed might fit together more consistently with a bit of tinkering.

IV. Going Farther - New Options and Connections.

Having re-walked the previously surveyed terrain, we can turn now to explore a
few additional features of the logical geography of critical options. A good point to begin
is with the empiricist assumption mentioned above to which supporters of the argument
might appeal on behalf of their reply to Q1. They claim that Mary would learn something
new when released, no matter how much physical knowledge she might have had
before. That confidence in turn derives largely from the complementary empiricist view
that her lack of pre-release red experience guarantees her lack of certain kinds of
experiential knowledge. According to what we may call “The Pineapple Principle,” in
allusion to Locke’s original example of the taste of that then exotic fruit, some
knowledge is essentially empathetic or subjective and can be gained only by undergoing
the relevant experience. As Locke writes in the Essay on Human Understanding
(1689),

I think it will be granted easily that if a child were kept in a place where
he never saw any other than black and white till he were a man, he would
have no more ideas of scarlet or green, than he that from his childhood
never tasted an oyster, or a pine-apple, would have of those relishes.

(Book II, Chapter 1, Sect 6.)

The Pineapple Principle, or something like it, seems to drive the intuition that Mary’s prior knowledge would inevitably be incomplete, even though Jackson himself denies that the Knowledge Argument relies on it. (1998) The relevant assumption is not quite the same as Locke’s original, since he was concerned with the necessity of sensation as the means of acquiring certain ideas. Were someone to have induced a red experience in Mary before her release through direct cortical stimulation rather than by holding a red object before her eyes, that would suffice to end her isolation and void the antecedent conditions of the argument’s main hypothetical. What is crucial about Mary before her release, is not merely that she has never seen actual red objects, but that she has never had a red experience produced in any way, whether by red visual stimuli or otherwise. What matters is the range of her past experiences, and the relevant empiricist assumption concerns the limits on what we can know or understand about experiences of kinds that we have never had. If you have never had a taste of pineapple experience, then there are some real respects in which you lack knowledge about it or about what it is like.

One further qualification may be needed for those who accept some form of physical supervenience: i.e. the principle that two beings alike in every physical respect share all their other properties as well. If one accepts such supervenience, a counterexample to the pineapple principle might arise if we could imagine constructing a molecular duplicate of someone, call her Ramy, who has had past red experiences but is not having any now. (Unger 1966) Since the duplicate DRamy is newly created it has itself never had a red experience. Yet since DRamy is a molecule-for-molecule duplicate of Ramy and Ramy knows what red experiences are like, it would follow by the supervenience thesis that DRamy has such knowledge as well. Thus given
supervenience, the possibility of such duplicates would seem to deny the strict necessity of actual past experience as a cause of present knowledge of the relevant type of experience.

Thus if one wants to formulate the empiricist assumption in a way likely to command the widest agreement it may be necessary to accommodate such cases, but they pose no real problems for the advocates of the knowledge argument. Since its supporters accept its dualist conclusion, many of them might likely reject supervenience as well. Alternatively if their aim is to construct an argument whose steps will be accepted by those who come to it as physicalists, then a suitably qualified version might be constructed that should still suffice to underwrite the intuition they wish to support. Thus the argument is not affected either way by whether or not molecular duplicates might share a color-experienced person’s knowledge of what it’s like. Mary is not such a duplicate, and the question is whether or not she lacks such knowledge before her release. All the formulations of the Pineapple Principle, whether qualified or not, would seem to indicate she must. Thus the contest of intuitions moves to the empiricist principle itself: are there forms of knowledge about experience that can be had only by those who have themselves had such experiences (or their duplicates)?

The truth or falsity of the principle is most directly relevant to Q1. Supporters of the knowledge argument might buttress their intuition that Mary learns something upon her release by appeal to what the principle tells us about what she could not have known earlier. The appeal is not mandated; some supporters of the argument might reject the Pineapple Principle but still give a positive answer to Q1. One could do so consistently, but at the cost of needing to find other support for one’s intuitions about Q1 and Mary’s gain of knowledge. Though the Pineapple Principle seems most directly relevant to Q1, its overall implications for the Knowledge Argument are less obvious though perhaps no less important. As we will see below the principle, or at least a near
relative of it framed in terms of the *capacity to have* specific types of experience, may be not only consistent with physicalism but entailed by some plausible versions of it.

The view that lies behind the Pineapple Principle is largely the same as that which leads Thomas Nagel to his claim that it is impossible for humans to fully understand what it is like to have bat-like echo-locatory perceptual experiences (Nagel 1974). We lack the phenomenal modalities to undergo or project ourselves fully into alien forms of experience that lie outside our phenomenal range. We are thus limited in our ability to know or understand all the subjective phenomenal features that one might come to know or understand from the internal perspective of those who do include them in their experiential repertoire. To state the matter more in Nagel’s terms, there are *subjective facts* about experience, where “subjective” means a fact that can be fully understood only from a specific perspective, namely that associated with the ability to have or undergo such experiences oneself. Only a creature able itself to experience red can fully know or understand the subjective fact of what it is like to have a red experience. If full understanding or knowledge requires empathy, then one can have such projective understanding only for states that are enough like those that we ourselves can be in. In so far as we can not be experientially like a bat, we have a parallel lack in our ability to know or understand it is like to be conscious in that bat-like way. More specifically, in so far as we can not *be* experientially bat-like, we can not fully *understand or know* what it is like to experience bat-like.

Cases like the bat seem to show that there are subjective facts and empathetic ways of knowing that give access to facts that would not otherwise be open to understanding. Some physicalists might challenge the empiricist assumption for reasons like those that lead them to answer “No” to Q1. They might call into doubt our negative intuitions about what Mary supposedly could not know, by reminding us of the absolutely comprehensive nature of the physical knowledge attributed to her by
stipulation. Can the argument’s proponents assume that, even given such knowledge on a scale and level so unlike any that we come close to having at present, Mary still would not know or understand the facts about what seeing red is like? Is the intuition that the argument requires about such a radically contrary to fact case so obvious that it precludes reasonable doubt to the contrary? My sympathies lie with the Pineapple Principle and the belief in subjective facts, but I must confess that the opposing intuition can not be readily dismissed as obviously mistaken.

However, the more interesting question, is what implications the existence of subjective facts might have for the knowledge argument or for the state of physicalism. Both Nagel (1974) and supporters of the knowledge argument seem to regard subjective facts as incompatible with the truth of physicalism. If one is thinking of physical knowledge as solely third person knowledge of the sort one might get from studying physical theory, then physicalism’s claim to completeness might seem to preclude there being any subjective facts knowable only from a given experiential perspective. Physical science facts are objective facts par excellence; they are of just the sort that might be understood from a plurality of perspectives, e.g. by us and by Martian scientists who do not share our sensory modes but who nonetheless accept similar physical theories. Thus if “physical facts” means “physical theory facts”, then subjective facts - if any exist - would seem to fall outside that range.

However, if one reads “physical facts” in a broader way to mean all the facts “that obtain in virtue of physical processes” or “that are realized by underlying physical structures”, then it is far less clear that no such facts might be subjective in the sense of being perspectivally restricted in their knowability. Nonreductive physicalists in particular will be quick to deny that any such exclusion is entailed. Ontologically reductive physicalism, i.e. the metaphysical view that everything real is physically realized, does not entail representationally reductive physicalism or require that the
resources of physical theory should suffice for representing and understanding everything that can be known or understood about any feature of the physically realized world. As noted above, nonreductive physicalism rejects conceptual-representational imperialism and the view that physical science gives us all the tools we need for understanding in every context of application, whether it be ecology, economics or psychology. 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 understandingly engage their own physically realized dynamic features in ways that cannot be duplicated by any cognitively equivalent external mode of access.

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 here in detail, 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 their 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 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 as promised above 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 that 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.

P1 stipulates a conjunction that might embody a hidden contradiction. 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. P1 may be intended as a stipulation, but if it ends up specifying a contradiction or a conjunction of conditions that can not be jointly satisfied, then we are justified in rejecting it whether stipulation or not. We can thus add rejecting P1 on such grounds to our “Take your pick” list of options for resisting the argument. That makes five and counting.

The teleo-pragmatic aspect of self-understanding systems provides the basis as well for the other sort of amendment to the survey that I promised above, that of showing further links among the prior identified options which may end up blurring them a bit, by showing ways in which they overlap or interdepend. Q2, Q3 and Q4 gave critics the options of restricting Mary’s gain respectively to know how, to new ways of knowing old propositions, or to new but merely fine grained propositions. The three
seem to offer distinctive progressively more restricted points at which to draw the line against the argument, and there is some validity to that view.

However, matters get a bit more confused if we push deeper into the details and implications of any of the three. Those, like Lewis, who wish to draw the line at Q2 will claim that Mary gains only new abilities. But the relevant abilities are highly cognitive in nature and provide mastery of new concepts of the sort required for understanding new fine grained propositions. In so far as our grasp of concepts and ability to deploy them in thought depend upon practical intra-mental skills, real cases in which we gain such skills will be ones as well in which we come to know new fine grained propositions, ones which we previously not only failed to know but which we could not even have understood or grasped. Thus there is more in common than might at first have seemed by between those who aim to draw the line at Q2 by limiting Mary’s gain to know-how and those who part company at Q4 by crediting her with knowledge of new propositions, but only in the fine grained sense.

The situation is similar with respect to Q3. Coming to know an old proposition in a new way will also typically involve both the acquisition of new skills and the mastery of new concepts. For example, if one gains a new mode of direct causal access to some feature of one’s mind, one will in the process also likely acquire new skills for recognition and engagement of the sort that would in turn pragmatically anchor one’s command of new introspective concepts and thus one’s ability to comprehend new fine grained propositions incorporating those concepts.

Thus what we had earlier identified as three separate options for disputing the intuitions behind the knowledge argument turn out to be at least partly interdependent. Does that fact weaken or strengthen the case against the argument? On one hand it might seem to do the former, in so far the argument is not forced to prevail in three distinct and completely independent conflicts of intuitions. If the argument’s success
required adopting its preferred alternative on three separate and controversial issues that might seem to stack the odds against it, but if those three could be collapsed it might seem more likely to succeed by winning a single clash of intuitions.

On the other hand, the mutually reinforcing links between the various options for disputing the argument might increase their individual plausibility against the competing intuitions required in each case by the argument. Perhaps more importantly, the three viewed together as a joint but multi-faceted option might more effectively explain away the intuition that Mary gains knowledge of new propositions in the coarse grained sense needed to secure its anti-physicalist conclusion. Any inclination to think of her as doing so might be explained away as merely her gaining new modes of access and associated new skills that give her the mastery of new concepts and thus the ability to comprehend new fine grained propositions. The unified option might thus provide a more effective counterweight against the epistemic intuitions on which the argument relies. Although the three choice points associated with drawing the line at Q2, Q3 and Q4 may not be fully independent, they nonetheless present three issues on which the argument’s intuitions must win against viable competitors, which are made all the more plausible by their mutual incorporation in a consistent alternative view of Mary’s epistemic change. In that respect the links between the three options may strengthen the overall case against the Knowledge Argument and provide a means to answer the disunity challenge raised at the end of section III, about which I will say more after considering one last line of reply.

Our final amendment focuses on the distinction between coarse and fine grained propositions and its implications for the argument. As Lewis (1988) notes, what is needed to defeat physicalism is the conclusion that Mary learns a new proposition in the coarse grained sense and that she thereby reduces her uncertainty and more narrowly locates herself in logical space. Anything less gets dismissed as mere redescription or
translation - like learning to say in Urdu what you were already able to say in English.

Your mastery of a new medium of representation in that sense produces no increase in your expressive power or your semantic information, defined as the elimination of epistemic possibilities that thus decreases uncertainty about the nature of the actual world and its location in the space of possible worlds. It may seem that if any matters of fact and actuality are left open in that sense to Mary’s state of knowledge until her release, then

- the physical story is incomplete,
- there are truths about the world that fall outside its scope, and
- physicalism is thus false.

However, we must be careful here because many physicalists view those as three different consequences, and nonreductive physicalists in particular would deny that the third follows from the first two. They agree that the physical science “story” is not the only valid representation that we need for our successful cognitive engagement with reality, and they also agree that there are true things that can be said, meant, and understood by us that fall outside the expressive range of what we, as the cognitively and causally situated agents that we are, can say within the resources of physical theory. The special sciences give us means of saying, understanding and knowing facts that are not accessible to us through the structure of physical theory.

However, the nonreductivist aims to hold onto her physicalist credentials in so far as everything real, including everything real described by the special sciences, is physically realized. It is this fact of the physical as the underlying realization base that supposedly anchors its claim of ontological physicalism. Even if there are complex regularities that we can not apprehend through the medium of physical theory but only through special sciences concepts and modes of access, they too as well are real in virtue of their underlying physical realizations. It would be pointless to try to do biology
in the language of microphysics, but that does not undermine the fact that all the properties of biological systems are realized in their underlying physical structure and organization.

The same holds true, when the alternative non-microphysical system of representation rather than being that of some higher level natural science is instead that embodied in the structure of reciprocal causal engagement that underlies our intra-mental self-understanding as conscious experiencing physical systems. The relevant modes of access and engagement give us ways of apprehending, recognizing and most importantly of dynamically interacting with our own mental nature that we could not possibly duplicate through the medium of physical theory. But that is consistent with both sides of the epistemic engagement being physical or physically realized.

Both the understander and what gets understood may owe their reality to the underlying physical structures that realize them, but it does not follow that having a microphysical description of those bases, even a complete microphysical description, would in itself enable us to understand all the real features that we are able to grasp through other systems of representation and engagement such as those associated with our reflexive self-awareness. According to the nonreductive physicalist, realization suffices to ground a claim of ontological physicalism without entailing any claims about the universal adequacy of physical theory as a means of representation and understanding.

Thus two crucial and interdependent issues arise. First in what sense are the relevant higher level facts and propositions new and different from those that concern the 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.

Thus 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 might seem to maximize the definiteness of one’s location and thus to preclude learning any new coarse propositions. However, one must be careful in moving from matters of metaphysics to epistemological conclusions, just as when one does the converse. In particular we should be cautious in moving from matters of realization and causal determination to conclusions about knowledge and understanding. Just because A-facts supervene on B-facts, it does not follow that knowing the B-facts suffices for knowing, or even for being able to understand, the A-facts. The aesthetic qualities of a painting may supervene on the point-by-point distribution of pigments on its surface, but knowing all the latter facts does not by itself confer knowledge or understanding of its aesthetic qualities.

Perhaps all one lacks in such a case is simply knowledge of some fine grained propositions and mastery of the concepts they embody. Once we know the full facts about the base, it may seem we could gain at most new ways of redescribing those
same facts; there may be no room to gain knowledge of new propositions in the coarse
grained sense of eliminating previously open possibilities.

That intuition has appeal, and in my earlier survey of the argument I did not
consider any challenge to it. However, it too may offer options for disagreement. One
could perhaps answer Q4 in the affirmative conceding that Mary learns a new coarse
grained proposition, and yet still deny that her doing so refutes physicalism. Finding
room for such an option requires care in thinking about what counts as knowing one’s
location in logical space. It’s a bit like the old cartoon joke in which a lost visitor
confronts an informational display which announces “You are here *,” but fails to provide
any further map to explain where “here**” is. One might similarly 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 located.
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 in so far 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
course 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. Knowing Euclid’s axioms does not suffice for being able to state or recognize all his theorems; mathematics otherwise would be much easier than it is. However, Mary’s problem is not so much a matter of determining which of a given set of propositions are true (or follow as true), but rather of simply being able to grasp or understand them. Consider the analogy with the mathematical case. One might both know the axioms and be able to state some conjectures that one understands, but yet be uncertain as to which if any of them are theorems until one had constructed the required proofs. However, one might be in an even worse epistemic state if one lacked some of the concepts needed to express the relevant propositions. Having mastered the concepts needed to state and understand the basic axioms of number theory does not guarantee that one also has a similar command of those needed to state all its theorems. A mathematical novice might command the elementary concepts of
arithmetic but lack any concept of imaginary numbers. Before one can know all the propositions that follow from the axioms one must at a minimum be capable of understanding them and having command of all the requisite concepts. Similarly to get from propositions about the realization base to those about realized higher order properties, one must have mastered the concepts needed for understanding those latter propositions.

If having a red experience is a higher-order physically realized state, then is there any reason for supposing that the concepts needed for understanding it must be within Mary’s grasp prior to her release? Her ability to understand all the propositions about its physical base would not seem suffice to guarantee her mastery of the requisite higher-order concepts. Indeed her lack is likely to be greater than in the mathematical case. Even if one initially lacks more advanced concepts of number theory, they are at least definable in terms of the elementary concepts that one must master to understand the axioms. Thus those further concepts are cognitively accessible from one’s initial epistemic situation even if one lacks practical command of them at the outset.

No similar guarantee need apply in the case of realizer and realized. The concepts needed to grasp propositions about higher-order realized facts need not be 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. (Van Gulick 1992) 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 an adequate 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 various underlying physical bases. As noted above, the function from worlds to truth values associated with the higher-order HP proposition differs from that associated with the proposition about any of its particular realizations RB. Moreover, to understand the HP proposition one must also grasp what it is that makes all the higher order cases similar despite their diversity of realizations. Indeed understanding the relevant structure of logical space will require not only that we understand the resemblances among the diversely realized instances of HP, but also the resemblances and differences between them and the equally diversely realized instances of other higher order, experiential properties HP₁, HP₂, HP₃, ........ Without a grasp of those aspects of the structure of the containing logical space, we could not understand the HP proposition nor how it figures in the structure of the space within which it allows us to locate ourselves. Lack of such a grasp would put us back in our “You are here*” cartoon predicament.

Thus it seems we have yet another and later option for disagreeing with the Knowledge Argument and resisting its anti-physicalist conclusion. We might concede that Mary post-release both learns new coarse grained propositions and comes to better understand her location in logical space, and yet deny that this refutes or is inconsistent
with ontological physicalism understood as the claim that everything real is physically realized. Indeed nonreductive physicalists may find it the most plausible point at which to draw the line.

A defender of the knowledge argument may object that any such reply relies upon shortchanging Mary on her physical knowledge, which by stipulation is complete. He may claim that I have shown at most that Mary might learn something upon release despite having earlier known all the propositions about the realization base and all the microphysical facts and laws. However, the Knowledge Argument credits Mary pre-release with complete physical knowledge simpliciter, without any restriction to base or microphysics. Such universal physical knowledge might be taken to include knowledge as well of all the physically realized facts and propositions. However, if real experiences are all physically realized as the physicalist claims, then it would seem Mary would have to have known those realized facts before as well. So there would not be anything left for her to learn, just as the original argument claims. All parties can agree on the following conditional:

IF experiences are physically realized and Mary’s knowing all the physical facts requires knowing all the physically realized facts, 

THEN she cannot learn any new coarse fact about experience upon release.

But the critic of the argument will charge the reply with begging the question in assuming that Mary could have known all the physically realized facts prior to release. If there are subjective physically realized facts - as nonreductive physicalism implies - then Mary could not know all the physically realized facts about experiencing red before herself having had the relevant experience.

Thus in a sense our two new critical options interconnect, even though one of them comes before our earlier four questions and the other afterward. The very earliest
question of whether it is possible for Mary to know all the physical facts and propositions before her release links up with the new final question of whether her coming to know a new proposition after release would refute physicalism. The answers in both cases may turn on how one defines the range of “physical facts” and on whether or not there are physical facts that are subjective in the sense of being open to understanding only from a restricted range of experiential perspectives. According to the nonreductive physicalist there are some facts that are physical in the sense of being physically realized but are nonetheless subjective. If so, then Mary pre-release could not know all the physical facts, if that is meant to include all the physically realized facts. Similarly Mary post-release might come to know a new proposition without refuting physicalism, as long as the fact she learns is both physically realized but not able to be adequately grasped or understood nonempathetically through the resources of physical theory.

It is here that the line begins to blur between the fine and coarse grained options for resisting the argument. The fourth resistance point in the original survey conceded that Mary learned only a new fine grained propositions, i.e. that she merely came to conceptualize an old fact in a new way but learned no new coarse proposition nor came to locate herself more narrowly in logical space. Our newly added final option turns as well on reconceptualization, but takes a more expansive view of what it involves or makes possible. Mary’s acquisition of introspectively anchored concepts on this latter option, enables her both to understand a new and previously ungraspable propositions and to articulate the structure of logical space in a way that gives her a better understanding of her actual location in it. The two options offer their respective ways of looking at what Mary gains through acquiring such experiential concepts. I tend to prefer the latter, but whichever way one opts the original anti-physicalist conclusion is blocked.

V. A New and Expanded Map of Options.
Let me sum up where things stand. The four critical options identified in the original survey all remain. They may be more interdependent that earlier suggested, but that may strengthen rather than weaken the overall case against the knowledge argument. Two new options have been added, one coming before the progression in the original survey and the other coming afterward. We can thus add two new questions to our diagnostic tree to locate these additional ways of parting company with the argument. Let us call them Q0 and Q5 to indicate the way in which they bracket our original Q1 - Q4.

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

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

Both questions turn on matters of consistency. Q0 asks whether it is consistent for Mary to have never experienced red but yet to know all the physical (or physically realized facts) about such experiences. The believer in subjective physically realized facts will argue that the two conditions are not compossible. Q5 similarly asks whether or not physicalism is consistent with the existence of subjective facts, ones that can be learned or understood only by undergoing the relevant sorts of experience.

Critics of the argument who are nonreductive physicalists will argue that they are consistent; indeed nonreductive physicalists will say the existence of such facts is entailed by physicalism properly understood. The equation of the real with the physically realized need not conflict with Mary’s learning new propositions post-release as long as some physically realized facts are subjective.

We can summarize the newly expanded space of critical options for responding to the argument in an enlarged and improved diagnostic tree as in figure 2 below. Once again, I make the same invitation, “Choose your favorite. Take your pick.”
NO, not if there physical subjective facts.  YES
X  Physicalism not refuted.

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

NO (Churchland, Hardin)  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.  Gains new knowledge of propositions.
(Lewis, Nemirow)  YES
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.
(Tye, Horgan, Churchland)  YES
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.
(Lycan, Loar)  YES
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?

NO (Nonreductive Physicalist)  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

What of the disunity concern raised by the editors of this volume (Stoljar and Nagasawa)? As noted at the end of section III, the very diversity of responses to the
Knowledge Argument might be taken to show that physicalists have failed to find a clear answer to its challenge. If they cannot agree about where or how it goes wrong, can they be confident that it does so? This is a serious objection, but as we have just seen the various replies may be less in conflict than they seemed in our original survey. With a bit of tinkering, I think there are plausible ways to consistently combine key elements from multiple replies, especially if one takes a teleo-pragmatic view of the problem. For example, one might argue that prior to release Mary cannot know all the physically realized facts, because her lack of practical intra-mental abilities prevents her from grasping certain concepts and thus of understanding fine-grained propositions involving those concepts. Indeed her practical limits may even prevent her from comprehending or articulating the structure of logical space in a way that is adequate for understanding some coarse-grained propositions about red experience. There may be not only concepts but properties, physically realized higher-order properties, to which she has no adequate cognitive access prior to release. Unlike Lewis and other supporters of the ability reply, the teleopragmatist does not claim Mary lacks only know how. Rather the claim is that because she lacks the requisite practical abilities she lacks the ability to grasp various fine or even coarse grained propositions about physically realized facts, and thus she could not have complete physical knowledge in her pre-release state, if doing so required knowing all the physically realized facts. In so far as one can produce such hybrid replies, perhaps I should amend my invitation: not merely, “Take your pick,” but “Take more than one if you like, as long as you do the needed tinkering.”

VI  Jackson’s Rejection.

As a final question we might ask where we should now locate Frank Jackson in our tree, given his rejection of his own past argument (1998) and his reasons for doing
so which he explains in this volume. Where might those reasons fit in our taxonomy? Does his recent change of mind align with any of the six options we have considered? The answer is both yes and no. He embraces the ability reply and like Lewis and Nemirow views Mary’s epistemic gain upon release as merely a gain in know how. So he might be listed with them as giving a “only know how” reply to Q2. But his reasons for doing so and for denying that Mary lacks any prior propositional knowledge turn crucially on his acceptance of the representational theory of perceptual experience, according to which the phenomenal character of a red experience is exhausted by its representing the world as containing objects with the objective property of redness. Red experiences are not red, and in particular they do not have any property of phenomenal redness (PHRED-ness). Indeed Jackson holds there is no such property as PHRED, and thus Mary cannot be denied access to it prior to her release, nor are there any true propositions about experiences having PHRED that she fails to know or grasp. The property her experience has is not that of being phenomenally red (PHRED) but merely that of phenomenally representing external objects as red. And that latter property Jackson holds is one that Mary could at least in principle deduce from her prior comprehensive physical knowledge. Thus he also aligns with those, like Churchland, who think Mary could infer all there is to know about red experience from her physical knowledge. He fails to give a “No” to Q1 only because he regards Mary as gaining know how.

Jackson attributes his prior acceptance of the knowledge argument to what he now regards as the illusory belief in the property of phenomenal redness (PHRED-ness), and he prescribes the representational theory as the necessary means for dispelling the illusion. Thus although he ends at the same point as Lewis and Nemirow, he arrives there by an interestingly different route, and thus he enriches the critical options.
However, I remain skeptical about his claim that one can defeat the Knowledge Argument only by accepting the representational theory. Given the diversity of critical alternatives we have surveyed, it is far from obvious that representationalism offers the only viable line of reply. Moreover, I am uneasy about holding rejection of the Knowledge Argument hostage to the truth of the representational theory, which I am not alone in finding less compelling than does Jackson. (Block 1990) To make his solution work Jackson needs to sail a narrow course between two perils. On one hand he must explicate what he means by the representational nature of an experiential state in a way that is rich enough to make plausible his claim that it exhausts its phenomenal character. It must for example include much more that just the state’s satisfaction conditions, but also something of how it phenomenally represents those conditions. And what of inverted spectrum cases? If they are possible, should they count as states that are phenomenally different though alike in how they represent the world as being? On the other hand, Jackson must give an account austere enough to support his claim that an experiential state’s representational nature can be fully inferred from the physical facts by pre-release Mary. Perhaps he can sail that course, but I think it remains as yet an open question. Jackson himself does not attempt to prove the truth of the representational theory. He defers to other advocates and treats it more as an assumption in offering his diagnosis of where the Knowledge Argument goes wrong. Nor can I here at the end of this chapter try to disprove the theory. I must content myself with hoping to have shown that those who may have doubts about it still have plausible alternatives for answering the Knowledge Argument. Please add Jackson’s option to the mix, but if you have qualms about the representational theory, you still have lots of others from which to pick.

References


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


Jackson, Frank (200x) “Mind and illusion” This volume, xxx-yy.


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


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


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.
