Philosophia
https://doi.org/10.1007/s11406-024-00717-5

What Counts as Cheating? Deducibility, Imagination, and the Mary Case

Amy Kind[1]

Received: 2 January 2024 / Revised: 2 January 2024 / Accepted: 5 January 2024
© The Author(s) 2024

Abstract
In *The Matter of Consciousness*, in the course of his extended discussion and defense of Frank Jackson's famous knowledge argument, Torin Alter dismisses some objections on the grounds that they are cases of cheating. Though some opponents of the knowledge argument offer various scenarios in which Mary might come to know what seeing red is like while still in the room, Alter argues that the proposed scenarios are irrelevant. In his view, the Mary case is offered to defend the claim that phenomenal facts cannot be *a priori* deduced from physical facts. Thus, a proposed scenario constitutes an objection to the knowledge argument only if it presents a case in which Mary's learning inside the room comes about via *a priori* deduction from physical facts. Call this *the deducibility standard*. In what follows, I'll explore a series of relevant cases in an effort to clarify this standard. Doing so enables us to better understand how cheating should be assessed in this context and thereby also to get clearer on the argumentative dialectic surrounding the Mary case.

Keywords Knowledge argument · Deducibility · Imagination · Phenomenal consciousness

1 Introduction

Since the launch of Chat-GPT in late November, 2022, those of us who grade student work have had to confront a whole new set of questions about what counts as cheating. Though many of us take it as a firm starting point that a student who submits a paper that was wholly generated by AI has cheated, there are other cases where the matter less clear cut. Suppose the student writes an initial outline entirely on their

✉ Amy Kind
akind@cmc.edu

[1] Department of Philosophy, Claremont McKenna College, Claremont, CA, USA

Published online: 18 January 2024

own, but then the actual paper is written by ChatGPT on the basis of the inputted outline? Or suppose the student asks ChatGPT to help it come up with some objections to the argument under consideration, or to help it find illustrative quotes to bolster their own argument? What if a student writes an initial draft of their paper entirely on their own, but then uses ChatGPT to proofread their work for grammar and style?

In the face of some of these questions, we might start to wonder whether our notion of cheating needs to be revised. In particular, it starts to look like cheating can only be assessed relative to a specified standard, i.e., that what counts as cheating depends entirely on how an instructor frames the assignment. If we take this view, however, a further question instantly arises, for we might wonder whether a given instructor's specified standard is an appropriate one to which students should be held.

An analogous set of questions about cheating were very much on my mind as I read Torin Alter's excellent book *The Matter of Consciousness*. In offering an extended discussion and defense of Frank Jackson's famous knowledge argument, Alter dismisses some objections on the grounds that they are cases of cheating. Though some opponents of the knowledge argument offer various scenarios in which Mary might come to know what seeing red is like while still in the room, Alter argues that the proposed scenarios are irrelevant. In his view, the Mary case is offered to defend the claim that phenomenal facts cannot be *a priori* deduced from physical facts. Thus, a proposed scenario constitutes an objection to the knowledge argument only if it presents a case in which Mary's learning inside the room comes about via *a priori* deduction from physical facts. Call this *the deducibility standard*.

With respect to some of the possible scenarios that an opponent might propose, I agree completely with Alter's assessment of foul play. Suppose, for example, that someone was to argue that while inside the room Mary could come to know phenomenal redness by pricking herself with a pin and having an experience of her (red) blood. To my mind, Alter is exactly right when he notes that this kind of example would be "plainly beside the point of the Mary case … It is cheating" (Alter, 2023, p. 168). But there are other possible scenarios where the matter strikes me as considerably less clear. In what follows, I'll explore a series of relevant cases in an effort to clarify how cheating is best understood in this context and thereby to get clearer on the argumentative dialectic surrounding the Mary case.

2 The Deducibility Standard

In reconstructing the knowledge argument, Alter sees it as proceeding in three stages. In the first stage, Jackson uses the Mary case to establish the existence of an epistemic gap between the physical and the phenomenal domains. In the second stage, Jackson argues that this epistemic gap entails the existence of a modal gap between these two domains. Finally, in the third stage, Jackson argues that the modal gap entails the existence of an ontological gap between these two domains. If there is such an ontological gap, then it follows that physicalism is false. My interest here is in the first stage of the argument, though as we will see, Alter's interpretation of the argumentative moves involved in the first stage depend on his interpretation of the

argumentative moves involved in the second stage. I will have nothing to say about the third stage of the argument.

So how exactly does the first stage of the argument work? Let's work through Alter's construal of it. As the Mary case is described, inside her black and white room, Mary is able to learn the complete physical truth. When Jackson himself explains what the complete physical truth consists in, he includes not just the truths belonging to physics but also those belonging to chemistry and neurophysiology. Yet as Alter points out, this specification seems arbitrary; if we are going to include truths of chemistry and neurophysiology among the physical, it is unclear why we wouldn't also include truths of microbiology or human anatomy (Alter, 2023, p. 19). Alter thus suggests we adopt a restrictive interpretation of what is meant by the "complete physical truth" such that it is understood in terms of the information belonging to fundamental physics. Because Mary has perfect powers of deduction, and because the truths of chemistry and neurophysiology (as well as microbiology and human anatomy) are all *a priori* deducible from the information belonging to fundamental physics, Mary would know all those truths as well. Indeed, she knows all the truths that are *a priori* deducible from the complete physical truth. (Going forward, in line with Alter's own practice, I will drop the qualifier and use "deducible" as shorthand for "*a priori* deducible.") According to what Alter calls *the learning claim*, when Mary leaves the room she learns new truths about what it is like to see in color. If the learning claim is true, and there are phenomenal truths that Mary does not come to know until she leaves the room, then there is an epistemic gap between the physical and the phenomenal.

On Alter's construal of the knowledge argument, then, the notion of deduction plays a key role in the establishment of the epistemic gap. This role becomes especially clear when Alter considers what he calls the *no-experience-necessary objection* (NEN). According to proponents of NEN, the learning claim should be rejected in light of its reliance on the false assumption that one must have color experiences in order to know what it is like to see in color. In response, Alter argues that NEN "is based on a misunderstanding" in that it "misses the crucial role that deducibility plays in the knowledge argument." (Alter, 2023, p. 55) On his view, in order to justify the rejection of the learning claim, it is not enough to show that Mary can in some way use the physical knowledge that she has inside the room to learn new phenomenal truths about color even though she has never had color experiences. Rather, one must show something more specific: The way that she uses the physical knowledge to learn the phenomenal facts must consist in a deductive process. This gives us the deducibility standard mentioned above.

Let's look at Alter's treatment of one defense of NEN, that offered by Dennett (2007), to see the deducibility standard in action.[1] Dennett asks us to consider Robo-Mary, a robotic version of Mary. RoboMary's hardware is equipped for color vision but she was brought online without it; she has only black and white cameras installed. However, in virtue of her vast knowledge, she writes a program that allows her to colorize the input from her black-and-white cameras, such that the grayscale pixelated

[1] Alter also considers a different version of NEN offered by Mandik (2009). I do not have the space to discuss that version here.

frames coming from those cameras are automatically replaced with colored frames. It seems plausible that RoboMary knows what color experiences are like.

After presenting the scenario, Dennett asks: "Is this a cheat or isn't it?" (Dennett, 2007, p. 28). Alter answers in the affirmative. What RoboMary does violates the deducibility standard. She hasn't used the physical facts to deduce what phenomenal facts are like; rather, she has used the physical facts to cobble together a workaround such that she manages to have a phenomenal experience even though she lacks color cameras. On this assessment, the RoboMary case doesn't establish that one can have knowledge of what color experiences are like without color experience (i.e., it doesn't establish that no color experience is necessary) but rather just presents us with a deviant way of coming to have color experience. There is thus no threat to the learning claim.

Anticipating this kind of response, Dennett offers a second case to support NEN: Locked RoboMary. Not only does Locked RoboMary lack color cameras, but she also lacks color vision. Software has been installed to make sure that her visual system is limited to grayscale values. But Locked RoboMary is very clever. She builds a model of herself without the software restriction. Though she is in state A when viewing a ripe tomato, she determines that the model is in state B when viewing the same tomato. By noting all the differences between state A and state B, she makes the necessary adjustments to her system and goes into state B. As Dennett claims:

> State B is, by definition, *not* an illicit state of color experience [she cannot have color experience because her color vision is locked]; it is the state that such an illicit state of color experience normally causes … But now she can know just what it is like for her to see a red tomato, because she has managed to put herself into just such a dispositional state. (Dennett, 2007, p. 28)

Confronted with this case, one might question whether Locked RoboMary really knows what it is like to see red. One might worry that the dispositional nature of Locked RoboMary's knowledge disqualifies it from being genuine phenomenal knowledge. But Alter is happy to grant Dennett the conclusion that, as a result of bringing herself to be in state B, Locked RoboMary does come to know what seeing red is like. Even so, in Alter's assessment, the Locked RoboMary case fares no better than the original RoboMary case when it comes to the deducibility standard. Just as RoboMary failed to come by her phenomenal knowledge by way of deduction, Locked RoboMary too fails to come by her phenomenal knowledge by way of deduction. Here again, the phenomenal knowledge is gained by cobbling together a workaround. Granted, this time the workaround does not produce a phenomenal experience but only the dispositional state that one would be in were one having a phenomenal experience. But that doesn't change the basic logic of the situation: Because Locked RoboMary gains the phenomenal knowledge in virtue of being in state B, and not in virtue of deducing such knowledge from the physical facts, there is no threat to the learning claim. Like RoboMary, Locked RoboMary too is a case of cheating – it is "hardly better than if she had simply unlocked her color-vision system" (Alter, 2023, p. 60).

To my mind, Alter is clearly correct in his assessment that these cases violate the deducibility standard. Neither RoboMary nor Locked RoboMary deduce their phenomenal knowledge from the physical facts. Dennett seems to grant this as well. But he does not think this fact invalidates the force of the cases. When he considers the accusation that they involve cheating, he responds by calling the deducibility standard into question: "I just don't see that this [deduction from the physical facts] is what matters." In elaborating this point, he notes that the accusation of cheating "presupposes an improbable and extravagant distinction between (pure?) deduction and other varieties of knowledgeable self-enlightenment" (Dennett, 2007, p. 29). What should we make of this response? As we will see, the issue of what counts as pure deduction proves to be important.

3 Cheating and the Deducibility Standard

When Alter takes up Dennett's response, he confesses puzzlement (Alter, 2023, p. 61). This puzzlement seems justified. First, it is not clear what Dennett means by self-enlightenment – a notion that seems to have connotations irrelevant to the discussion at hand. But suppose we interpret "varieties of self-enlightenment" simply to be picking out non-deductive methods for coming to know phenomenal truths. In that case, the distinction Dennett is calling into question would be the one between deductive and non-deductive methods of attaining knowledge. And it is not at all clear why this distinction should be viewed as either improbable or extravagant. Dennett does little to help us understand the accusation.[2]

Dennett being Dennett, one can be forgiven for thinking that the descriptors in the accusation were chosen largely for heightened rhetorical effect. So perhaps the accusation simply boils down to the claim that the distinction is problematic and irrelevant.[3] Let's take these in reverse order.

To my mind, the charge of irrelevance clearly has little bite. Alter's discussion of the overall argumentative strategy of the knowledge argument makes clear why the notion of deducibility is needed. Recall that the second stage of the argument involves a move from an epistemic gap (there are truths about consciousness that cannot be deduced from the complete physical truth) to a modal gap (there are truths about consciousness that are not necessitated by the complete physical truth). The move from the epistemic gap to the modal gap is defended by way of a claim about the nature of the relationship between deducibility and necessitation, specifically, that it is impossible for X to necessitate Y if Y is not deducible from X (see Alter, 2023, pp. 16–18, 62). It's thus crucial to this move that the epistemic gap be expressed in terms of non-deducibility. An opponent of the knowledge argument might want to question Jackson's interpretation of the relationship between deducibility and necessitation, but such a question arises precisely because deducibility plays a crucial role

[2] After introducing the accusation, Dennett goes on to discuss a different variant of the Mary case that he calls Rosemary. To my mind, this case is not helpful in showing why the accusation holds, and I don't have the space to discuss it here. But see Alter's discussion (2023, pp. 61–62).

[3] This reading is in line with how Alter ends up interpreting the matter (see 2023, pp. 61–62).

in the knowledge argument. The distinction between deductive and non-deductive ways of attaining knowledge is thus undeniably relevant to the Mary case.

So we can dismiss Dennett's charge of irrelevance. But what about his charge that the distinction is problematic? On the one hand, the notion of deduction on which Alter is relying seems uncontentious. As Alter makes clear in offering some introductory remarks on terminology, he uses "deduction" to mean "deductive reasoning that does not depend for its justification on sense experience, as when one does a proof in logic or mathematics" (Alter, 2023, p. 5). But while this looks to be a fairly standard way of cashing out the notion of deduction, it leaves open some important questions that are directly relevant for the matter at hand. In short, I worry that there may be some ways that Mary might attain knowledge of what color experience is like that do not fit neatly within this conception of deduction but that also do not seem so easily dismissed as instances of cheating. In the remainder of this section, I'll pursue this worry. Though I'll take no stand on whether this is what Dennett had in mind in criticizing Alter for using a problematic ("extravagant") distinction, I think it's plausible to take his inclusion of the parenthetical "(pure?)" when referring to deduction in the context of the deductive/non-deductive distinction as a gesture in this direction.

To start, consider the diagrams that often accompany mathematical proofs in various areas of mathematics ranging from geometry to analysis. Consider, for example, the many diagrams appearing in Euclid's *Elements*. What is the role of such diagrams? Baker (2020, § 2.1.3) usefully summarizes three possible scenarios: (1) The diagrams are wholly dispensable. They are mere visual aids and play no role in the logical structure of the proofs. (2) The diagrams are logically dispensable but psychologically indispensable. While as a practical matter it is difficult or even impossible to understand the proofs without the diagrams, they play no role in the logical structure of the proofs. (3) The diagrams are not only psychologically indispensable but also logically indispensable. The diagrams play an essential role in the logical structure of the proofs. Consider, for example, Jesse Norman's view that a diagrammatic logic has "informational resources in virtue of its representational form which are not available to a logically equivalent sentential counterpart" (Norman, 1999). Once we properly understand the role that diagrams play in logical inference, we end up with a view of inference that "is not straightforwardly 'deductivist,' but incorporates observational and experimental processes" (Norman, 1999).

Let's suppose that some of Euclid's diagrams fall into the third scenario just delineated. In that case:

> deleting all the diagrams will render many of the proofs invalid. This raises the further question of whether a distinctively diagrammatic form of reasoning can be identified and analyzed, and—if so—whether it can be captured in a purely deductive system. (Baker, 2020, § 2.1.3)

Diagrammatic reasoning thus seems to problematize the distinction between deductive/non-deductive ways of knowing. It seems to present us with what seems like an in-between case. To go back to Dennett's use of the parenthetical "(pure?)": One can see how there might be a temptation to view diagrammatic reasoning as deductive reasoning but not *pure* deductive reasoning.

Of course, Alter could deny that Euler's diagrams fall into scenario (3). But that would require him to take a substantive position on a disputed matter in philosophy of mathematics. Alternatively, perhaps he could deny that diagrammatic reasoning counts as deduction. But that would require him to deny that Euclid's *Elements* serves as a paradigm of deduction – and, as Baker notes, Euclid's *Elements* is "often held up as a canonical example of the deductive method" (2020, § 2.1.3). Moreover, this would be a difficult position for Alter to take in light of his employment of what he calls *trapezoid case 1*: a budding geometer proves the sum of a trapezoid's angles by constructing a proof from Euclid's axioms and postulates. In Alter's view, trapezoid case 1 is a clear example of deductive reasoning (Alter, 2023, p. 61).

So, let's continue with the assumption that diagrammatic reasoning is not a case of pure deductive reasoning. Now suppose we wanted to hold a geometer to some kind of geometric deducibility standard. Would the use of diagrams violate it? Should it count as cheating? To my mind, it's hard to swallow the suggestion that a geometer who uses diagrams is engaging in foul play. The use of diagrams seems wholly different, for example, from the clear cheating involved in what Alter calls *trapezoid case 2*. In that case (simplifying slightly), a budding geometer time travels to the future and acquires a device that puts a person in the state that someone is in when they know the sum of the trapezoid's angles, as long as the person is contemplating the axioms and postulates when the device is activated. As Alter notes, when the geometer uses the device, even though the axioms and postulates play an essential role in the process they used in coming to know the sum, "that process is plainly not deduction" (Alter, 2023, p. 61).

Where does that leave the defender of the deducibility standard? It seems to me they could go two ways here. They could grant that there's no violation of the deducibility standard in Euler's *Elements* and count diagrammatic reasoning as deductive, perhaps by expanding the notion of deduction to include impure deduction. Or they could refuse to count impure deduction as deduction, and thus treat the use of diagrammatic reasoning as a violation of the deducibility standard, but they could couple this with a denial that this kind of violation of the standard amounts to cheating. In other words, they might divide violations of the deducibility standard into two classes: harmless violations and harmful (cheating) violations.

Unfortunately, however, whichever option they choose they are going to run into trouble when we return to Mary. To see why, it will be helpful for us to consider a geometer with very strong powers of imagination. They don't need to draw the diagram on paper; rather, they simply imagine the diagram in their mind. They use the imagined diagram just as they would use the physically-drawn diagram; the imagined diagram plays an essential (scenario 3) role in their proof construction. But again, as is the case with the geometers who use drawn diagrams, there's no foul. Either this use of imagination fails to violate the deducibility standard or it is a harmless violation. Using the imagination isn't cheating.

So let's consider ImaginautMary, who has spent a lot of her spare time inside her black and white room practicing her imaginations skills. When ImaginautMary uses her imagination inside the room in an effort to determine what seeing red is like, there can likewise be no foul. Imagination is not off-limits. If ImaginautMary succeeds in her endeavors – if she is able to leverage the physical facts that she learns while

inside her black and white room to jump start her imagination, and if by way of her imaginings of redness, she comes to know what seeing red is like – then she would come by this knowledge without having experience of red. So this gives us a new version of the no-experience-necessary objection and, importantly, it is one that cannot be dismissed for misunderstanding the role that deducibility plays in the knowledge argument. Just as with the case of the geometers, the way that ImaginautMary uses her physical knowledge to learn the phenomenal facts cannot be dismissed by invoking the deducibility standard (either there's no violation, or the violation is a harmless one). More generally, the deducibility standard no longer seems as powerful a weapon in fending off possible objections to the learning claim.

4 Conclusion

Importantly, the considerations of the previous section do not themselves shed any light on whether Mary actually can imagine color from her room. But it's perhaps worth noting that Dennett, for one, seems to think that it's a live possibility:

> We are told that Mary in her cell can't imagine what it's like to experience red, try as she might. But suppose she doesn't accept this limitation and does try her best, cogitating for hours on end, and one day she tells us she just got lucky and succeeded. "Hey," she says, "I was just daydreaming, and I stumbled across what it's like to see red, and, of course, once I noticed what I was doing I tested my imagination against everything I knew, and I confirmed that I had, indeed, imagined what it's like to see red!" (Dennett, 2007, p. 23)[4]

Paul Churchland has long pursued a similar line. On his view, Mary's neuroscientific knowledge ensures that she has the relevant neuroscientific concept for the phenomenal state a person is in when seeing red. This neuroscientific concept will be quite sophisticated, for example, it might be specified in terms of "various spiking frequencies in the nth layer of the occipital cortex (or whatever)." Thus, even though Mary has never been in the relevant cortical state, Churchland thinks that she may well have the ability to imagine being in it, and to do so with "substantial success." (Churchland, 1985, pp. 25–6) At the very least, Churchland suggests that *it's not unthinkable* that she has this ability, that her having this skill is not "beyond all possibility" (Churchland, 1985, p. 26).

Speaking for myself, while I am inclined to deny that pre-release Mary has the ability to imagine red, I am also inclined to deny that her having this ability is unthinkable. Moreover, as I have argued elsewhere, reflection on what skilled imagiers can and cannot do – and in particular, reflection on how one can use a process of imaginative scaffolding to build one's way from experiences that one has had to imagine experiences that one has not had – makes the matter a bit muddier than is often thought (Kind, 2019, 2020). At the very least, I worry that too often discussion

[4] For related discussion, see Dennett (1991, p. 399).

about the Mary case simply relies on a knee-jerk skepticism about the imaginative capacities that Mary has prior to her release.

But we do not need to settle the question of whether pre-release Mary can imagine red for the case of ImaginautMary to have bite. In raising the case, my purpose is not to show that the learning claim is false. Rather, I use the case to help highlight some of the limitations of the deducibility standard. As I have tried to show in this commentary, this standard is not as cut-and-dried as it initially appeared, and this unclarity complicates the argumentative dialectic. Unfortunately for proponents of the knowledge argument, when it comes to the issue of cheating, matters concerning the Mary case turn out to be tricky as matters concerning ChatGPT. In both domains, it turns out to be surprisingly hard to specify in a principled way what the boundaries of cheating should be.

Acknowledgements Thanks to Robert Howell and Jan Michel for helpful feedback on an earlier draft.

Funding No funding was received to assist with the preparation of this manuscript.
Open access funding provided by SCELC, Statewide California Electronic Library Consortium

Declarations

Competing Interests The author has no relevant financial or non-financial interests to disclose. The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alter, T. (2023). *The Matter of Consciousness*. Oxford University Press.

Baker, A. (2020). Non-reductive methods in mathematics. In Edward Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). Retrieved from https://plato.stanford.edu/archives/sum2020/entries/mathematics-nondeductive/.

Churchland, P. (1985). Reduction, qualia, and the direct introspection of brain states. *Journal of Philosophy*, *82*, 8–28.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown.

Dennett, D. (2007). What RoboMary knows. In Torin Alter & Sven Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 15–31). Oxford University Press.

Kind, A. (2019). Mary's powers of imagination. In Sam Coleman (Ed.), *The Knowledge Argument* (pp. 161–179). Cambridge University Press.

Kind, A. (2020). The skill of imagination. In C. Pavese (Ed.), *Routledge Handbook of Skill and Expertise* (pp. 335–346). Routledge.

Mandik, P. (2009). The neurophilosophy of subjectivity. In John Bickle (Ed.), *Oxford Handbook of Philosophy and Neuroscience* (pp. 601–618). Oxford University Press.

Norman, J. (1999). Diagrammatic reasoning and propositional logic. London: University College London. Retrieved from https://www.proquest.com/docview/1775113388 via ProQuest Dissertations Publishing.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.