

The Thought Experiment

Mary is imprisoned in a black-and-white room. She learns about the outside world only through books printed in black and white, a black-and-white computer, and a black-and-white television screen. Mary learns all the relevant facts about color vision that can be learned in this way. She becomes an expert not only in neuroscience but also in physics and chemistry, and thus gains theoretical knowledge about all the relevant facts involved in color vision. She knows, for example, that what we see as red is caused by the excitement of long-wavelength cones in our eyes, and how the brain translates this into red. One day, Mary is released from her cell, and she sees the color red for the first time. Now it seems if physicalism is true—that is, if it is true that all our knowledge can be reduced to knowledge about physical facts—that Mary learns nothing new when she sees a red flower just outside the prison door. But Mary does learn something new. From this, we can conclude that physicalism is false.

Source of the thought experiment:

Jackson, Frank. 1986. "What Mary didn't know." *The Journal of Philosophy* 83, no. 5: 291–295.

Reflection by Martina Fürst, University of Graz

Think of the first time you had a particular experience: the first time you tasted an exotic fruit, felt snow on your skin, or smelled the ocean. Maybe some of your friends or family had already enjoyed this kind of experience, they told you about it, and they tried to describe it—nevertheless, having the experience was something completely different. No description, comparison, or information about the causes of the experience conveyed to you what you gained by having the experience yourself: knowledge of what the experience is like, of the experience's so-called phenomenal character.

The thought that there is something special about experiences, namely, that their phenomenal character cannot be conveyed by descriptions, is already present in C. D. Broad (1925). He argued that even a logically omniscient archangel, endowed with the power to perceive the structures of atoms, could not find out that a substance with a particular structure has, for example, the distinctive smell of ammonia to humans. If Broad was right, then it is not our limited knowledge about physics or chemistry that explains why we cannot figure out what an experience is like before having it. Rather, theoretical descriptions are *fundamentally* unable to convey what experiences are like. Jackson's thought experiment about Mary the scientist elaborates in detail the idea that no theoretical knowledge suffices for knowing the phenomenal character of experiences—not even complete physical knowledge (broadly understood to encompass neurophysiological, chemical, and other knowledge). Notably, the thought experiment does not only reveal that *knowledge* about experiences is special. It also aims to show that the very *nature* of experiences, and of consciousness in general, is special.

The nature of consciousness and how it relates to the body is one of the fundamental questions in the philosophy of the mind. In the past century, many philosophers aimed at reducing consciousness to a physical phenomenon. For instance, on one view, conscious states such as

red experiences are just brain states. On another view, experiences are states that play a particular role (e.g., of detecting red objects in the environment), and various physical systems (e.g., brains, computers) can realize this role. What is common to these accounts is that consciousness is ultimately physical and can be explained in physical terms. But wait: didn't we just see that no physical theory can tell us what an experience is like?

Mary's room is based on this insight, but it aims higher: at elucidating the very *nature* of experiences. To reach this aim, Jackson assumes that Mary knows *all* physical facts; that is, he ascribes knowledge of a complete physical theory to a subject. This combination highlights perfectly the puzzling relation between objective theory and experiencing subject. On the one hand, Mary knows all the physical facts about color experiences, but, on the other hand, she still cannot deduce what the color experiences will be like. Therefore, some facts about consciousness—in particular, facts about what experiences are like—are not physical facts, Jackson concludes.

The intuition that physically omniscient Mary learns something when having her first color experience is compelling. Some philosophers (for example, Daniel Dennett) reply that this intuition is misleading, that Mary learns nothing. Others take the intuition seriously, but reject the conclusion about the nonphysical nature of consciousness. They ask: What is it that Mary learns? Does she really learn new *facts* about color experiences? Accordingly, alternative explanations of what Mary learns have been proposed that do not entail the nonphysical nature of experiences: for example, David Lewis and Lawrence Nemirow think that Mary just acquires a new capacity, such as knowing how to recognize red experiences. Others, for example, Katalin Balog, Terence Horgan, and Brian Loar argue that Mary relearns a fact she already knew, but in a new way: in her achromatic room, she grasped facts about red experiences from an objective perspective, and now she grasps the very same facts, but from her subjective perspective.

On my view, the latter explanation is particularly interesting, since it elucidates a feature of consciousness that makes its investigation very special. With regard to consciousness, we have two ways of accessing it: from our inner, subjective perspective, and from a third-person, objective perspective. But which of these two perspectives reveals the nature of experiences?

I think Jackson is right in that enjoying an experience from the subjective perspective reveals something significant about this experience. In particular, I suggest the following analysis of experiences: what makes an experience the very experience it is, is its phenomenal character. If so, the subjective perspective directly reveals *essential* properties of experiences (see also David Chalmers, Martine Nida-Rümelin). Accordingly, theories of consciousness that are stated in objective, physical terms cannot adequately account for experiences, since they have to leave out facts about the essential properties of experiences; facts about what experiences are like. Recall again the experience of tasting an exotic fruit for the first time, an experience that until then you only knew by description. On my view, by having this experience, you did not just grasp the experience in a new way. Rather, you gained direct insight into an essential property of this experience—an insight that you could not have gained in any other way. If this analysis is correct, then the nature of experiences, and more generally, of consciousness is not (entirely) physical.

The question of whether science ever can account for what it is like to have an experience is known as the “puzzle of consciousness.” Mary's room helps us to understand more deeply this philosophical puzzle. Thinking through various analyses of the scenario opens the door to possible solutions to one of the deepest puzzles in the philosophy of mind.

Discussion Questions

1. Think of the first time you had a particular experience. Prior to having it, could others have conveyed to you what the experience would be like? What, if anything, changed when you enjoyed the experience yourself for the first time?
2. Do you think that (future) science will explain consciousness just as well as other (for example, biological or physical) phenomena, or do you rather think that consciousness is a special topic of investigation? If so, in what sense is consciousness special, and what would a full explanation of it require?

Further Reading

Alter, Torin, and Robert J. Howell. 2008. *A dialogue about consciousness*. Oxford: Oxford University Press.

Broad, C. D. 1925. *The mind and its place in nature*. London: Kegan Paul.

Coleman, Sam, ed. 2019. *The knowledge argument*. Cambridge: Cambridge University Press.

Fürst, Martina. 2014. "A dualist account of phenomenal concepts." In *Contemporary dualism: A defense*, edited by Andrea Lavazza and Howard Robinson, 112–136. New York: Routledge.

Nida-Rümelin, Martine, and Donnchadh O Conaill. 2019. "Qualia: The knowledge argument." In *The Stanford encyclopedia of philosophy*, Winter 2019 edition, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2019/entries/qualia-knowledge/>.