For *Notre Dame Philosophical Reviews*, published June 12.

Stalnaker, Robert C., *Our Knowledge of the Internal World*, Oxford University Press, 2008, 149pp., $39.95 (hbk), ISBN 0199545995

Clare Batty, University of Kentucky

Robert Stalnaker's *Our Knowledge of the Internal World* is an extremely engaging and rewarding book. Originally presented as the John Locke Lectures at Oxford University in 2007, it covers a broad range of issues related to our knowledge of thoughts and phenomenal experience—our own and others'. Its discussion contains a familiar cast of characters: Gustav Lauben and Rudolf Lingens, Ralph and Ortcutt, O'Leary and Daniels, Kripke's Pierre, Sleeping Beauty and Lewis' two gods. But the central protagonist of the book is Jackson's Mary. His focus is not on what Mary can tell us about physicalism per se but on what Mary can tell us about our epistemic situation with respect to our own thoughts and phenomenal experiences. The book's discussion is both dense and concise, at times requiring the reader to closely examine the text multiple times or to fill in details not explicitly stated. But, even when particularly challenging or tricky, the book more than rewards its reader by challenging her to tackle lingering problems about the mind and knowledge with typical 'Stalnakerian' precision and rigor. Any philosopher will be better off for having read the book.

Driving the book is Stalnaker's commitment to externalism and to a model of content first introduced in "Assertion" (1978). And while its arguments aren't anything that would convince the internalist to convert to externalism, this is not Stalnaker's chief aim. Rather, his main purpose in the book is to advance his position by showing that his model can accommodate, or explain away, allegedly internalist-friendly problems such as the knowledge argument and others relating to self-locating belief.

In his opening chapter, Stalnaker introduces us to the topic and methodology of the book. He tells us: "My subject matter will be that part of our knowledge that the Cartesian internalist takes to be the most basic and unproblematic—knowledge of our own phenomenal experience and thought. But I will approach the subjective point of view from the outside" (2). He spotlights four debates in which there has been a general shift towards externalism, but the lesson of the chapter lies in how to understand the externalist project. Stalnaker stressed that it is a mistake to understand the externalist's "approach [to] the subjective point of view from the outside" as one that is perspective-free—a 'fully objective' view from nowhere. The externalist's project is conducted from *within* the world and, although this is to admit that the theorist's project necessarily involves a 'view from somewhere', this needn't detract from the absoluteness of the conception that follows from it. This is a theme to which Stalnaker returns throughout the book.

Having clarified his position 'in' the subject matter, Stalnaker turns to Mary. As we know, Mary is a brilliant neuroscientist who is confined to a black and white room. By reading black and white books, Mary has come to know all of the physical facts about color vision. The question: when she is let out of her room, and shown a ripe tomato, will she learn anything? Jackson says 'yes'; Mary learns what it is like to see red. Therefore, physicalism is false; there are non-physical facts. In chapter two, Stalnaker rejects the three standard approaches of dissolving the problem: the Fregean strategy, Lewis' and Nemirow's ability hypothesis and Perry's self-location analogy. Both the Fregean strategy and the ability hypothesis fail because they cannot avoid the conclusion that Mary's ignorance consists in her lack of information, in her inability to rule out possibilities. But this is the very claim they deny. Perry's self-location analogy comes the closest to what we

need to dissolve the problem. But it doesn't quite do it. Perry's notion of reflexive content provides the resources with which to claim that Mary gains new information but it cannot do so for the point after her release at which Mary allegedly acquires the information that threatens physicalism.

Still, Stalnaker claims that the analogy with self-locating knowledge points us in the right direction. According to Stalnaker, Perry's view is attractive because it recognizes the importance of a kind of information that involves the subject and her place in the world. In chapter three, Stalnaker develops his own account of self-locating knowledge with the aim of developing the analogy with phenomenal knowledge. He proposes a modified Lewisian account that places greater emphasis on the connection between the subject who has the beliefs and the way that the subject takes the world to be while, at the same time, allowing that those beliefs distinguish between real possibilities. According to Stalnaker, his model of self-locating belief makes explicit the relation between an absolute conception of the world and a conception of a subject's perspective on it.

In chapter four, Stalnaker compares this case to the case of phenomenal knowledge. Stalnaker argues that what the case of self-locating knowledge shows us is that there may be distinctions between possibilities that can only be represented by a subject from within a certain context. (This he shows through an analysis of Lewis' two gods and Elga's Sleeping Beauty case.) Stalnaker calls the information imparted in contexts of this kind *essentially contextual information*. What makes the case of self-locating knowledge analogous to phenomenal knowledge is that fact that both are essentially contextual. In each case, the knowledge gained is knowledge that could only be acquired in a certain context. If this is true, then the information that Mary acquires when she leaves her room is information that she could not have had while inside her room. Still, before she leaves her room, Mary knows all the information there is to know about the world in itself. I will focus on this aspect of Stalnaker's discussion in what follows but, before I do, let me give a brief synopsis of the remaining three chapters.

Stalnaker recognizes that his argument that phenomenal knowledge is essentially contextual undermines the idea that we have the kind of direct contact with our minds that internalism claims we do. On a modified Mary example that he employs in chapter four (also quoted below), Mary cannot tell if she is having an experience of red or an experience of green—even though she is currently undergoing one of them (an experience of red, as it turns out). She knows that she is having an experience of red or green, but she does not know which one it is. Stalnaker's account of Mary, then, places her in a second predicament. She gets out of her room only for us to discover that she doesn't know what kind of experience she is having! In chapter five, he follows Lewis in rejecting the idea that we are directly acquainted with our phenomenal experiences. With this goes the special, foundational, epistemic status that internalism claims experience has—a move he claims Lewis failed to make. This special epistemic role allotted to the mental, Stalnaker reminds us, has been a sticking point in the internalism/externalism debate and, in chapter six, he tackles the problem of epistemic transparency. Using Boghossian as his example, Stalnaker argues that internalists have mischaracterized externalism. The solution to the alleged problem of transparency, he argues, is a thoroughgoing externalism and a robust dose of contextualism about knowledge. We shouldn't think of our thoughts as internal sentences that we have infallible access to. If we grant that the content of thought is individuated by things external to the mind, then externalism is in trouble. But this *isn't* externalism, Stalnaker argues. It is, as he claims, "essentially an internalist picture with an externalist component grafted onto it" (130). The way we should

think of the content of our thoughts is as essentially *attributed*. Theorists attribute thoughts to thinkers as a means of explaining their rationality and behavior. And sometimes the thought that we attribute to others depends on the context that we are in. A modified epistemic transparency is preserved on this view given the aim of such attribution—namely, to appropriately explain the rationality and behavior of a thinker. If that's true (which Stalnaker claims internalists such as Boghossian claim it is), then we can say a thinker knows the content of her thoughts insofar as a theorist's description of them reflects the world as it seems to that thinker.

Earlier I stated that there are parts of Stalnaker's book that are particularly challenging. In what follows, I want to draw attention to a stage of the book at which the discussion is particularly tricky and potentially confusing. This is the discussion of chapter four in which Stalnaker's insights about perspective and the absolute conception come together with those about self-locating knowledge in an assessment of Mary's epistemic situation. Earlier in chapter two, Stalnaker draws attention to the form of Jackson's knowledge argument:

P1. Mary knows all the facts of kind K.
P2. Mary does not know the fact that P.
P3. So, the fact that P is not a fact of kind K.

Later, at the beginning of chapter four, he tells us that "[t]he debates about Mary and the knowledge argument raised questions about the extent to which features of our representations of certain facts (facts about phenomenal experience, "what it is like") belong to a conception of the world as it is in itself (to what Bernard Williams called an "absolute conception")" (75). They also raise questions about what it would take for such an absolute conception to be complete. Jackson thinks that it would require more than just a representation of the K facts (the physical facts); we would also need to add a representation of the phenomenal facts (the P1-P3 argument forms part of his argument for this conclusion). Given this background, I take it we can view Jackson's argument as containing a further conclusion:

P4. So, P facts enter into an absolute conception of the world.

One could dissolve the argument in four ways. One could:

i. deny (P1),
ii. deny (P2),
iii. deny the inference to (P3), or
iv. grant (P1)-(P3) but deny the inference to (P4)

I take it that Stalnaker opts for (iv). Consider the modified Mary story that Stalnaker presents in chapter four:

> Suppose that Mary, still in her room, is told that she will be subjected to the following experiment. She will be shown either a red or a green star, to be chosen by the flip of a coin, and she is told in great detail the exact circumstances of the two possible scenarios. So given her extensive knowledge of neurophysiology and color science, she knows that when the experiment is performed, she will be in the presence of a star with one of two specific light reflectance properties, and will be in one of two specific brain states. Both before and after the experiment is performed, there are two possible worlds compatible with Mary's knowledge—call them worlds R and G. As it happens, the red star is chosen, so she is in fact in possible world R. (86)

Quite rightly, Stalnaker follows this story by claiming that something changes about Mary's epistemic situation when she is shown the red star. He concludes that "you can explain the knowledge Mary lacks when she is still in her room, and you can understand that knowledge in terms of the elimination of possibilities without invoking the hypothesis of phenomenal information" (87). Mary knows all of the objective facts (P1). There is nothing else that we could have told her in her room that could have led her to this information (P2). So, Mary

learns something new but something different in kind (P3). Given the analogy with self-locating knowledge, the information that Mary acquires is essentially contextual. In order to get it, she needed to come out of her room and look at the star. In particular, he claims that Mary acquires the ability to represent information about *this* experience. But, like self-locating knowledge, this kind of information is not information about the world in itself (contra P4). Still, according to Stalnaker, the information she acquires distinguishes between real possibilities—between ways the world is and ways the world could have been—and in such a way that one does not have to appeal to special phenomenal information to explain her newfound abilities. In gaining the ability to represent information about *this* experience, Mary is able to eliminate possibilities—although Stalnaker is not explicit about which possibilities she is thus able to rule out.

Although I take this to be Stalnaker's official line on his modified Mary case, it is clouded by some discussion that occurs in an earlier section of chapter four (which, in turn, is presented as summarizing the analogy with self-locating knowledge). Stalnaker recognizes that the analogy with self-locating knowledge might seem strained. But, given some of the discussion in this earlier part of the chapter, it is also confusing just what the analogy with self-locating knowledge is supposed to *be*. In particular, some of the early discussion of the chapter encourages reading Stalnaker as opting for (ii) above. In this earlier discussion, he seems to indicate that what changes about Mary's epistemic situation is not her ability to eliminate epistemic alternatives but rather her ability to represent possibilities that she could only previously describe. So, at the very least (and I suspect also at the very worst), Stalnaker's lead-up to his assessment of Mary is puzzling.

Leading up to this assessment, he summarizes his model of self-locating knowledge with a familiar theme: "[t]he model recognizes that any conception of the world is necessarily a conception that is formed from a certain place in the world, using the materials that are available then and there, but this fact does not prevent the conception from being an absolute conception, in the sense that its *content* is concerned exclusively with the way the world is in itself" (78). Indeed, he goes on: "[o]ne of the things that emerged from the discussion of self-location is that there may be distinctions between the possibilities (the ways the world might be) that can be represented only from a certain perspective" (78). He then considers the Sleeping Beauty case. According to Stalnaker, when Sleeping Beauty awakens on Monday, she is able to represent in thought a distinction between two possible worlds: one in which *today* is Monday and one in which *today* is Tuesday. On Sunday, she was merely able to describe these two worlds that she could not distinguish. Given what comes immediately before the discussion of Sleeping Beauty, it seems consistent with what Stalnaker says here that, on Monday, Sleeping Beauty acquires the ability to represent possibilities that existed (and that she was able to acknowledge existed) before she went to sleep, but that she does not acquire any new information (contra P2). Rather, she is able to represent the 'same old possibilities' in a way she couldn't have before, and that it is this ability to represent, as opposed to rule out, possibilities that is essentially contextual. This is a much weaker claim than Stalnaker's official claim that, on Monday, she was able to eliminate possibilities (and in doing so acquires new information) that she was unable to previously.

Now Mary's situation is supposed to be analogous to that of self-locating knowledge in an important way. And the 'coin-toss Mary' case appears open to the same kind of reading. Before Mary leaves her room, she is able to describe two possibilities that she cannot distinguish—namely, the possibility that she is in the R world and the possibility that is in the G world. But when she leaves the room and is shown the star, she can,

at this point, represent possibilities that existed (and that she was able to acknowledge existed) before she left her room. As Stalnaker indicates, she might express these possibilities in the following way: 'Either *this* is an experience of red or *this* is an experience of green. I just don't know which it is'. (As I indicated earlier, it is this claim that leads Stalnaker to the view that Mary, and indeed all of us, do not have the kind of acquaintance with our experiences that the internalist takes we do.) And, indeed, Stalnaker tells us (as quoted above) that "[b]oth before and after the experiment is performed, there are two possible worlds compatible with Mary's knowledge—call them worlds R and G" (86). On the proposed analogy with Sleeping Beauty, then, what happens to Mary is that she is now able to represent in thought something that she could previously only describe. This proposal is in keeping with her situation being essentially contextual; but rather than any knowledge being essentially contextual, it is merely her representation of 'the same old possibilities' that is so (contra P2). Again, this is much weaker than Stalnaker's official claim that, upon her release, Mary is able to rule out certain ways the world might be.

No doubt Stalnaker could provide relief for this puzzlement by fleshing out how these newfound representational capacities are connected to the ability to rule out additional possibilities. Until he has done this, it is difficult for us to see exactly how to differentiate his view from other available options. And, no doubt, in the typical 'Stalnaker-effect', the richness of his insights will seep in over time and in such a way that it will become obvious that he has said so much more that you thought he had. (As it has with me. But one has review deadlines to meet.) But such is the beauty of his work. It is challenging, rigorous and innovative—and it continues to make you think long after you have closed the cover.