Précis of William S. Robinson’s *Epiphenomenal Mind: An Integrated Outlook on Sensations, Beliefs, and Pleasure* (New York and London: Routledge, 2019).

The central theme of this book is that the three divisions of our mentality indicated in the subtitle share the following properties. They are dependent on our brains, and they are not rightly thought of as causal contributors to what happens in our brains or in our behavior. The reasons for accepting these claims, however, are significantly different in the three divisions. Instead of there being one formulation of epiphenomenalism that strictly applies to all divisions, there is a family of arguments that lead to a similar conclusion in each case. It is this consilient family of views that form the integrated “outlook” of this book.

1. Sensations

The first four chapters of the book present a dualistic account of sensations. Unavoidably, part of the task of these chapters is negative. I explain why am not enamored of identity physicalism, functionalism, representationalism, higher-order thought theory, Russellian monism, Orchestrated Objective Reduction theory, or Integrated Information theory.

I cannot possibly go into all the reasoning in these critiques, but I will say a little more about one key argument. This argument is based on a compelling observation, namely, that our sensations give us no hint of the complexity of either their external physical causes or their brain event causes. I call this fact “SELL” – the Silence of Experience with respect to Lower Level items. (For example, for vision, color experience gives no hint either of anything remotely as complex as reflectance profiles, or as remotely complex as patterns of neural firings in V4.)

Dualists have an easy explanation of SELL. Complex external events cause complex neural events which in turn cause relatively simple, nonphysical events that are our experiences. In general, properties in effects can differ from properties of causes -- for example, hardness in a boiled egg is a different property from the heat that caused the hardening. It should thus not be surprising that a property in an experiential effect differs from both the causally relevant property of its neural event cause and the distal property that is distinctly causally relevant to that neural event.

Physicalists, however, face a number of unhappy choices. Here, I will briefly elaborate on just three leading ideas. (a) They may allow that instances of qualia are (in one of several ways) constituents of experiences, and hold that qualia are identical with certain complex physical properties. But then they will have no explanation of SELL. (b) They may deny the existence of experiences altogether, and swallow the irony that a leading criticism of epiphenomenalism – that it denies the allegedly obvious fact of causation of behavior by our experiences – applies in spades to their view. (c) They may hold that experiences represent qualities that are never instantiated in our world (although they might be instantiated in some possible worlds). They thereby incur the burden – unbearable in my view – of showing how a set of extensional relations can relate a physical event to a property that is nowhere actually instantiated, and thus can never be a property in virtue of which any causal contribution is ever made to anything in our world.

These negative explanations are required, but the main point of the first four chapters is to develop the dualism that it seems we must accept. A central question concerns causal relations. Here (as will be no surprise), I hold that brain events cause nonphysical sensations and that those sensations do not in turn cause anything. The brain events that cause sensations do, of course, causally contribute to further brain events. In general, events that causally contribute to brain events are either earlier brain events or stimulations of sense organs.

I also offer a theory of why changes in brains cause sensations. This theory turns on a fact related to SELL: The properties instantiated in sensations are, in a certain sense, simpler, relative to the properties involved in their causation. This fact suggests a conservation principle, where the conserved quantity is complexity-and-simplicity (CAS). On this view, there is a balance, or trade-off between the relative complexity of causes of sensations and the relative simplicity of the sensations that are the effects.

I do not claim to have proved this theory, and its acceptability depends on eventual neuroscientific discovery of an appropriate complexity property that would be distinctive of causes of sensations. I do claim that it is a coherent proposal, and that if it works out, it would be a solution to the Hard Problem.

I discuss the notion of transparency at some length, and argue that it is a mischaracterization of something else, namely, the three-dimensionality of (mainly visual, but also auditory and tactile) experience. This stance raises the question of what exactly we are doing when we attend to experiences as such. I argue that regarding an experience as an experience is an *addition* to experience, where the addition involves a pretense of an attitude of suspicion toward our experience.

2. Beliefs (and Desires)

According to the account of *Epiphenomenal Mind*, believing and desiring are global states of our organized brains. Our actions, both verbal and nonverbal, are symptoms of our brains’ organization.

The contrasting view that I reject holds that beliefs and desires are able to causally interact with each other (and with other beliefs and desires) to produce behavior.

A good way to understand the difference between these views begins with the observation that beliefs and desires are often referred to as (mental) *states*. These states are generally thought to be physical, or at least to supervene upon the physical. Either way, there should be a physical state corresponding to a mental state. We are thus entitled to ask what such physical states are states *of*.

In response, contemporaries will immediately think of *brain states*. I do not reject this answer, but I do ask whether this means (a) states of whole brains or (b) states of parts of brains. I think we have to accept (a). The picture that I reject adopts (b).

The key argument here is that no state of a part of a brain could determine what belief (or desire) it was supposed to be (or store, or represent, or correspond to). It does not matter whether we think of states of parts of brains as long-lasting patterns of connectivity among their sub-parts, or as ephemeral patterns of activation across the neurons composing a brain-part, or as some combination of such patterns. In any of these cases, the contribution of a part of a brain to behavior could be radically altered by its having a different set of neural connections to other brain parts. So, the fact that someone believes *p* logically cannot consist in some brain part’s being in state F. At best it could be the fact that part X is in state F *and part X has the set of connections {Y} to other brain parts* – where the set of connections is the connectivity to all of the rest of the brain. (A merely larger part will not do, since the same argument would apply: The effective contribution of a larger part would still depend on its connections to the remainder of the brain.)

This argument is shown to apply to contemporary discussions. The ‘P’s and ‘P\*’s in ubiquitous diagrams inspired by the exclusion arguments of Jaegwon Kim are commonly said to be physical states, presumably brain states; and thus they must be either states of whole brains or states of parts of brains.

To reject the picture presupposed by the use of Kim-style diagrams is to reject the view that beliefs and desires are the kinds of things that can causally interact to produce behavior. One might try to preserve mental causation by allowing that the global brain state that embodies our believing something can interact with the global brain state that embodies our desiring something. But the state of the whole brain cannot interact with itself. To say so would amount to no more than affirming the truism that our behavior proceeds from stimulation falling upon an organized brain.

We have thus arrived both at epiphenomenalism and the reason why its formulation must be a little different from epiphenomenalism for sensations. Sensations are events; thus they are of the right category to be causes. Epiphenomenalism says they exist, all right, they just don’t cause anything. But I do not say that beliefs, as they are thought of in the beliefs-as-states-of-parts-of-brains picture exist but are inefficacious. I say that beliefs and desires conceived of according to *that* picture don’t exist at all. Our behavior is caused by stimulation of our sense organs and the causal interactions of parts of our brains, but no part of our brain or state of a part of our brain is a belief, or the storer of or representation of a belief. These claims could be easily misunderstood if one abbreviated them to “Beliefs and desires don’t cause anything”, but I believe they will be recognized as what epiphenomenalism comes to when affirmed in conjunction with an accurate view of believing and desiring.

Now, there are of course many questions about believing and desiring that my rejection of the foregoing picture raises. I get at some of them by offering an account of what we are doing when we attribute a belief to others or to ourselves. I call this the “quasi-quotational account of belief attribution”. It says that attributing a belief that *p* to some person, S, is saying that S behaves like a person who says “*p*”. The usefulness of the practice of making such attributions is similar to the usefulness of direct quotation. If I tell you that S actually said “*p*”, and was not joking, disordered by drugs or disease, etc., you would form certain expectations about S. If I tell you that S believes *p* you would form a very similar set of expectations about S.

We can have such expectations about S even if S has never actually uttered “*p*”. There are many such cases. One example occurs when pronouns are used: S may have actually said only “He’s dishonest”, but in a context where it was clear that “he” referred to Jones. We can report this by saying that S believes Jones to be dishonest. Another sort of case occurs if we infer expectations about S from observing S’s nonverbal behavior. In that case, telling others that S believes *“p*” would be very like telling them that S had behaved as if S had said “*p*”.

The “expectations” here are, in general, not sets of explicitly formulated propositions about what S will do. They are tacit expectations; they are sets of dispositions to be surprised upon various possible future behaviors of S. But even with this clarification, the idea of behaving like a person who says “*p*” is likely to seem problematic. That is because of the well known fact that any belief is compatible with a vast range of behaviors, depending on what desires and what other beliefs people may hold, and any desire is compatible with a vast range of behaviors, depending on what beliefs and other desires one may have. My solution to this problem is that there can still be ways of behaving like a person who says “*p*” because there is an enormous number of beliefs and desires that people normally share, and that narrow the range of behaviors that are compatible with the whole set of a person’s beliefs and desires. An abnormal belief and desire can be held in some particular case, so our belief attributions are – of course! – subject to error. But when people are aware of unusual beliefs or desires, they generally include them in their reports of what others believe or desire. (Direct quotation can, of course, be misleading in the same way, but no one would conclude from that fact that it cannot be used to inform third parties about what to expect from a speaker.)

There are many concepts that are related to believing and desiring. Issues concerning some of these are conveniently explored by asking whether a robot could fall under them. I argue that possible robots could have intelligence, and (contra Searle) understanding. If they also had sensations – which I argue to be compatible with being a robot – they would have events that mattered to them (and not merely to their owners). I clarify, however, that the project of making a conscious robot (i.e., a robot that has sensations) is an entirely different project from making a robot that has intelligence or understanding, and that there is no reason to suppose that advances on the latter projects should make any advances on the former.

There are also two views that bedevil discussions of mentality, namely the view that we control our thoughts, and that our words *express* our thoughts. Control of our thoughts would involve trying to have them, which would require us to have already had them. This independent reason for rejecting control of our thoughts is another support for an epiphenomenalistic outlook.

Regarding expression (of thoughts), I argue that there is a sense in which “our words express our thoughts” is true. But the phrase is treacherous, because it suggests that our words must match a prior thought in order to be not just rambling. That suggestion just raises the question whether a prior thought must match a still earlier thought if *it* is not to be just rambling. The way out of the implicit regress is to allow that, while words would not mean anything if they were not products of an organized brain, we must not suppose that the process of producing our words involves a “matching” of words to anything that would count as a “thought”.

It is, of course, a consequence of this line of argument that the idea that our thoughts cause our behavior is a non-starter. An alternative view is that our brains produce highly organized inner speech, overt speech, and nonlinguistic behavior, and relations among these such that in large part (but obviously not always) our words and behavior “fit” (i.e., each makes sense when taken with the other). This view is another aspect of the epiphenomenalist outlook that this book advocates.

One might object here that my view goes so far as to imply – very implausibly – that all uses of “a thought” must be rejected as ill-founded, or as harboring confusion. However, I deny this implication. It is compatible with what I say that “a thought” can mean a proposition that one has thought of; which is to say, a proposition that one has stated in either overt or inner speech. “The thought that ‘*p*’ never occurred to me” can be adequately paraphrased as “I never uttered (either in written form, or in overt or inner speech) ‘*p*’ or any equivalent sentence.”.

3. Valence

Unlike sensations, pleasure and displeasure are directed upon aspects of our experience. They are also directly present in consciousness as reasons for the pursuit (or avoidance) of what is believed to cause what is found pleasant or unpleasant. That is, they have a phenomenology. This phenomenology, I argue, is caused by events in an evaluative system (ES), where an evaluative system is recognized by its role in promoting (or discouraging) the continuance or repetition of behavior that led to those events. As would be expected in an epiphenomenalist outlook, the phenomenology that is pleasure (or displeasure) is not efficacious, but, of course, events in the ES that cause valent phenomenology also have effects in the brain that lead to behavior.

This account of pleasure is defended against two other approaches that have some plausibility. It is then put to use in defending epiphenomenalism against one of the strongest arguments against it. This argument, which was advanced in 1890 by William James, turns on the idea that evolutionary explanations can operate only on what makes a causal contribution to behavior. So, if pleasure and displeasure were not efficacious, there would be no way to explain why we do not find necessary activities, e.g., breathing, to be painful, or detrimental activities, e.g. bruising ourselves, to be pleasant.

This argument seems plausible if we think of pleasure and displeasure as sensations that merely accompany other sensations and that, like sensations generally, can occur in any combination. But valence is not a sensation; it is the phenomenology of a type of evaluative-system event, where that type of event is identified by its causal contribution to behavior. This latter fact affords a place for natural selection to find a firm point of application.

4. Epilogue

The Epilogue addresses two main topics. The first is an objection to the effect that I have implausibly divided the mental into three different territories. A better theory, it is alleged, would give a more unified account, plausibly by regarding all aspects of the mental as representational.

My response to the specific alternative (i.e., representationalism) has been given at the beginning and at the end of the book, i.e., in Chs. 1 and 9, which offer reasons why representationalism is a bad theory of sensations and a bad theory of valence, respectively. My response to the general recommendation for a more unified view is that a proper account of the mental need not hold that all its aspects are of the same kind. An account is sufficiently integrated if distinctive aspects are adequately described and can stand in a coherent set of relations. I believe that readers of *Epiphenomenal Mind* will see that the differences among the aspects of mind that I discuss do not prevent them from living together in a coherent outlook.

It may, nonetheless, be objected that I have not given a reason why anyone should have gathered my three divisions under one concept. In response, I say that a fully adequate answer would require a history of the development of words such as “mind”, and I am not an historical linguist. However, there is a reason that I strongly suspect has a lot to do with the formation of our concept of mind. This is that all three of my divisions share a special relation to knowledge. To put the matter roughly, we do not find out about any of these matters by observation of the external world, and we are normally regarded as authoritative about such matters. We can, of course, lie about any of them. In the case of belief, we can find that we do not really know what we believe. Even so, however, declaring that one believes that *p* is (one way of) behaving like a person who says “*p*”. So, even if we are deceived about ourselves, there remains a close connection between what we are disposed to say we believe and what we believe that is unlike our relation to most declarations of what we believe and what those beliefs are about.

The second topic of the epilogue is a brief discussion of what an epiphenomenalist outlook means for our attitudes toward ourselves and others. One argument here is another aspect of the epiphenomenalist outlook; namely, that self-causation makes no sense. Another is that we must make, and continually bear in mind, a distinction that is somewhat difficult to keep in focus: namely, the distinction between responsibility for our actions (which we have) and responsibility for who we are when we do them (which, except in very special circumstances, we do not have). When we pay proper attention to this distinction, I argue, we will see that we are justified in praise, blame, reward and punishment for actions, but we should always have compassion for anyone who is in an unfortunate state.