Experimental Philosophy of Consciousness

Kevin Reuter

forthcoming in Oxford Studies in Experimental Philosophy, Volume 5 (Shaun Nichols & Joshua Knobe, eds.)

Abstract

Experimental philosophy of consciousness seeks to investigate and explain our thinking about phenomenally conscious states. Based on empirical studies, researchers have argued (a) that there is no folk concept of consciousness, (b) that we do not think Microsoft feels regret, (c) that unfelt pains are widely accepted, and (d) that people do not believe that duplicated hamsters have phenomenally conscious states. In this paper, I review these and other fascinating claims about people’s understanding of phenomenal consciousness. In doing so, I also show why experimental philosophy of consciousness is hard, although perhaps not quite as hard as studying phenomenal consciousness itself.

1 Introduction

Philosophers have often pointed out that there is nothing more familiar to people than their own phenomenally conscious states. These states include bodily sensations such as feelings of pain, emotions such as happiness and sadness, as well sensory experiences such as color experiences. Nonetheless, despite this sense of familiarity, reasoning about and understanding conscious states appear to be very difficult: Philosophers and laypeople alike find it difficult to make judgments about which other creatures have conscious states,

\footnote{The main topic of this article is \textit{phenomenal} consciousness. Thus, only states that are commonly considered to be something it is like to have or, in other words, that have a phenomenal, qualitative, or feeling aspect to them, will be discussed. Although moods form a further subcategory of phenomenal consciousness, they will not be discussed in this paper.}
whether there are degrees of consciousness, whether consciousness is unified, whether consciousness can be reduced to neuronal states, and so on.

Experimental philosophers of consciousness aim to understand how people understand consciousness; that is, they try to “meta-understand” consciousness. Of course, non-experimental philosophers have also paid significant attention to related questions, such as the meta-problem of consciousness (also see Section 4). However, the approach that experimental philosophy of consciousness takes is markedly different in at least two ways. First, experimental philosophers do not believe that the way experts understand a phenomenon such as consciousness is clearly representative of ordinary understanding. Accordingly, they believe that developing theories that are based on philosophers’ unchecked and possibly mistaken intuitions is problematic. Second, in order to obtain a more detailed, precise, and representative picture of people’s understanding of consciousness, experimental philosophers believe that it is necessary to collect empirical data using a wide variety of methods: The way in which people respond to questions about conscious states, the way that people talk about conscious states, and the way that people behave with regard to conscious states reveal insights into people’s understanding of conscious states that is almost impossible to attain from the armchair.

An important distinction needs to be made when studying phenomenal consciousness, as well as when studying people’s intuitions and understanding of phenomenal consciousness. We can either study consciousness by focusing on specific conscious states such as feelings of pain, emotions, and sensory experiences, or we can study consciousness as a general phenomenon that subsumes all types of phenomenally conscious states. Section 3 is devoted to the experimental-philosophical study of specific conscious states. Research in this area is strongly linked to existing theories. There is a simple reason for this: The concepts that are at the heart of philosophical theories of pain, emotions, and sensory experiences, are all part of our ordinary stock of concepts. Consequently, the corresponding folk concepts provide the starting point from which philosophical theorizing begins. Philosophical theories may deviate from our folk-conceptual understanding in order to make concepts more fruitful and precise (Carnap, 1950), but such deviations need to be justified. Given the importance of laypeople’s understanding of specific conscious states for philosophical theorizing, the importance of experimental-philosophical research in this regard is evident.

In Section 4, I review research on people’s understanding of conscious-
ness as a general phenomenon. Interestingly, research in this area proceeds more independently from philosophical and scientific theories of consciousness. Of course, it would be absurd to probe people’s intuitions about prominent theories of consciousness, such as Daheane’s (2013) global neuronal workspace model, or the integrated information theory by Tononi (2004). Thus, experimental-philosophical research on consciousness does not pretend to provide support for or falsify such theories of consciousness. Instead, experimental researchers focus on three related questions. First, do people have a concept of phenomenal consciousness and, if so, what does it look like? Second, do people make the same fundamental distinction between intentional and phenomenal mental states as experts? Third, to which creatures are people willing to attribute conscious states?

Although I would prefer to proceed directly to presenting and discussing experimental research on consciousness, I will begin in Section 2 with a brief analysis of the term “consciousness”. As will be shown, a study of people’s understanding of consciousness encounters some unique challenges. These challenges can be overcome, but they require a methodological approach that is quite different from what we find in other fields to which experimental philosophy has contributed. We will also see that these challenges have led to the development and application of exciting new methods.

I would like to make three qualifications. First, although I have attempted to present an impartial review, I will be assuming realism about phenomenally conscious states. Second, not every paper, and not even every influential paper in the field of experimental philosophy of consciousness will be discussed or mentioned in this article. Although this is unfortunate, it is also a positive indication that the field is growing quickly. Indeed, new areas of experimental philosophy of consciousness are constantly being explored, and the mutual influence between philosophers and cognitive scientists is increasing rapidly. Third, while this article mainly focuses on studies that investigate people’s beliefs, intuitions, and understanding of consciousness, experimental philosophers have also investigated aspects of phenomenal consciousness more directly. To mention just one example, Schwitzgebel (2007) empirically investigated whether people were constantly conscious of visual and tactile experiences (for example, in their left foot) during undisturbed moments in their lives. This was accomplished by having participants wear beepers and have them write down their experiences when these beepers made a sound. I consider such direct studies of phenomenal consciousness to belong just as much to the experimental philosophy of consciousness (widely
conceived). I have only adopted a narrower focus in this paper for reasons of space.

2 The Problem of a Missing Folk Term

Experimental philosophers typically investigate people’s intuitions and understanding of various phenomena by asking people questions; these questions are naturally couched in terms that refer to the phenomena in question. For example, when examining our understanding of knowledge in Gettier situations, epistemologists have asked themselves (if they are traditional philosophers) or have asked a large number of other people (if they are experimental philosophers) whether a person knows a certain piece of information when they are Gettiered.

It is essential for laypeople to have a good grasp of the terms that are examined in experimental studies. Whereas “know” is clearly a folk term, there is, unfortunately, no term for consciousness that is part of our folk terminology. Let us call this problem the problem of a missing folk term. One might initially feel comfortable dismissing this problem and following Carmel & Sprevak (2014), who state:

“We talk about consciousness in our everyday lives. We say that ‘she wasn’t conscious of the passing pedestrian’, that ‘he was knocked unconscious in the boxing ring’, that our ‘conscious experience’ of smelling a rose, making love, or hearing a symphony makes life worth living. Consciousness is what philosophers call a folk concept: a notion that has its home in, and is ingrained into, our everyday talk and interests.”

No! People do not talk like that. Philosophers talk like that. In most situations, people prefer to adhere to basic-level terms (see Rosch, 1978). At the basic level, we talk about joys and fears, feelings of pain and ticklishness, and experiencing colors and sounds. Superordinate categories such as emotions, bodily sensations, and sensory experiences subsume these states. Then there is consciousness, which includes emotions, bodily sensations, and sensory experiences, and is thus what we might want to call a superduperordinate category.

We do say things like “he was knocked unconscious”, but the meaning of “consciousness” in this case refers to being awake and responsive, and not to phenomenal consciousness.

2We do say things like “he was knocked unconscious”, but the meaning of “consciousness” in this case refers to being awake and responsive, and not to phenomenal consciousness.
The terms we use at the basic level are generally less likely to be technical (although there are of course exceptions), and thus more likely to be part of our stock of folk terms compared to terms at higher levels. If folk terms are primarily located at the basic level, we should observe that people use superordinate terms less frequently, which is indeed what we find. Table 1 below displays the frequency of terms in the Corpus of Contemporary American English (COCA) from three different domains of discourse for basic-level, superordinate, and superduperordinate categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Number</th>
<th>Term</th>
<th>Number</th>
<th>Term</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>superduperordinate</td>
<td>consciousness</td>
<td>21,011</td>
<td>artifact</td>
<td>3383</td>
<td>living being</td>
<td>354</td>
</tr>
<tr>
<td>superordinate</td>
<td>emotion</td>
<td>16,582</td>
<td>furniture</td>
<td>21,916</td>
<td>animal</td>
<td>58,376</td>
</tr>
<tr>
<td>basic-level</td>
<td>sadness</td>
<td>103,493</td>
<td>table</td>
<td>216,362</td>
<td>dog</td>
<td>98,847</td>
</tr>
<tr>
<td></td>
<td>anger</td>
<td>36,471</td>
<td>bed</td>
<td>114,879</td>
<td>horse</td>
<td>44,525</td>
</tr>
<tr>
<td></td>
<td>joy</td>
<td>34,712</td>
<td>chair</td>
<td>62,840</td>
<td>cat</td>
<td>41,763</td>
</tr>
</tbody>
</table>

Table 1: The frequencies with which terms are used in the Corpus of Contemporary American English (COCA) are displayed for three different domains (emotion, furniture, and animal) at different levels of categorization.

The relatively high frequency with which the term “consciousness” is used, namely 21,011 times, might be interpreted as contradicting the trend for lower numbers at higher levels of categorization. Such an interpretation is unwarranted for two reasons: (i) The majority of the uses have their sources in academic articles and scientific blogs (roughly 13,000), and (ii) the term “consciousness” is often used as synonymous with being awake, a notion in which we are not interested for our purposes.

Further evidence for the view that “consciousness” is not a folk term stems from cross-linguistic studies (in particular, see Étienne Balibar’s entry in Cassin et al.’s (2014) Dictionary of Untranslatables), which shows that consciousness does not have equivalent

---

3Categories at the superordinate level mainly serve the purpose of grouping elements for scientific or taxonomic purposes and are almost always theoretical constructs. Griffiths (2008) discussed the difficulty of developing a theory for a superordinate category such as emotion, and argued that three widely different types of states fell under the concept of emotion. Fehr & Russell (1984) conducted a series of experiments to investigate people’s thinking about both the superordinate category of emotion and the basic-level terms used to refer to states such as anger and joy.

4The adjective “conscious” (15,903 hits) is not more frequent than the noun “consciousness”.

---
terms in many other languages. A brief glance at the database *Childes*\(^5\) also reveals that children and adolescents hardly ever use such terms: There was only a single use by a nine-year-old child reading a story about a dog who checked whether a sick rabbit was still conscious. By contrast, the term “angry” was used 718 times, the term “hurt” was used 4348 times, and even the term “itch” was used 43 times by children.

If the term “consciousness” is no good, perhaps other terms can do the job (Chalmers, 2020, discussed several options). There are a few other candidate terms, but none constitute part of our folk vocabulary either. Terms such as “qualia” and “phenomenality” are obviously stipulative technical terms that are used to denote phenomenal aspects of conscious states. The term “experience” is a common term that is used by many people in everyday conversations, and is also used by experts to refer to phenomenally conscious states. However, as Sytsma & Fischer (md) convincingly showed in a corpus-analytic study, the meaning of the term “experience” in everyday conversations is distinct from the experts’ intended meaning. Furthermore, Wierzbicka’s (2010, 2019) cross-linguistic studies showed that “experience” was a construct of the English language, but not a “basic human” term\(^6\).

Where does this problem leave us? We cannot simply study people’s beliefs and intuitions about phenomenal consciousness in the same way that we study people’s beliefs about truth, beauty, knowledge, and responsibility. We need to find alternative ways to do so. The good news is that researchers have found such alternatives. The first strategy is to focus on a single type of conscious state for which both experts and laypeople possess folk terms. In Section 3, I will review the experimental literature on feelings of pain and color experiences, as well as emotional states, with a specific focus on happiness.

Researchers who adopt the second strategy, the extensional bottom-up strategy\(^7\), intend to study people’s intuitions about phenomenal consciousness more generally. In order to do so, researchers also ask participants about

---

\(^5\) *Childes* is a database featuring tens of thousands of conversations between children or adolescents and their caregivers.

\(^6\) The term “feeling” is perhaps the most promising candidate for a folk term, but it is limited because it cannot be applied to sensory experiences other than tactile experiences in a standard way.

\(^7\) The strategy is *extensional* because it investigate the states that are in the extension of the concept of consciousness, and is *bottom-up* because it makes inferences from the basic level of conscious states to the superduperordinate level of consciousness.
conscious states at the basic level. However, it is important to note that such researchers probe people’s intuitions about a variety of conscious states. Responses to questions about two or more mental states allow researchers to compare these responses and to determine whether laypeople implicitly group mental states according to the same categories that are defined by experts. In Section 4, we will discuss three of the central papers that have used the extensional bottom-up approach and the subsequent discussions that have ensued.

3 Experiences of Pain, Color, and Happiness

Color experiences, conscious pain and feelings of happiness are undoubtedly phenomenally conscious states. Many philosophers find the addition of “experience”, “feeling”, and “consciousness” to be superfluous, and claim that (in decreasing order of scholarly acceptance) “pain”, “happiness”, and “redness” already denote conscious experiences. However, such views are considered to be controversial, even within the expert groups. Whereas a majority of philosophers and scientists have argued that pain is identical to feelings of pain (Kripke, 1980; Aydede, 2006), only feeling theorists of emotions (James, 1884; Hufendiek, 2016) identify joy with a feeling of joy and anger with a feeling of anger (cognitivists strongly disagree with this identification; see Nussbaum, 2001). With regard to colors, most color experts do not identify colors with color experiences (but see Dennett, 1991, and Sytsma, 2010). Exacerbating the problem, we will shortly see that laypeople often use terms like “pain”, “color” and “happiness” in ways that differ significantly from philosophers’ definitions. I will now review some of the experimental studies of feelings of pain (Section 3.1.), color experiences (Section 3.2.), and emotions (Section 3.3.) before summarizing the main takeaways for the study of consciousness (Section 3.4.).

3.1 Feelings of Pain

The standard view among philosophers (as well as among pain scientists and medical experts) is that pain is a phenomenal state. In support of this view, it is often argued (Aydede, 2006; Hill, 2009) that pain has properties that are characteristic of phenomenal states: Pain is subjective and cannot
These two properties reflect two sides of the appearance-reality coin. If a state allows for an appearance-reality distinction, it is not itself a phenomenally conscious state. However, if a state does not allow for such a distinction, it is a phenomenally conscious state. With regard to pain, many experts consider that no appearance-reality distinction can be made: Pain and the feeling of pain necessarily go together:

If a person has a pain, then that person feels a pain. (Subjectivity)

If a person feels a pain, then that person has a pain. (No hallucination)

Neither of these statements is usually considered to require any further argumentation, but are claimed to be intuitively true. Furthermore, according to the standard view, these intuitions are not only available to those who have studied pain, but are said to be part of our folk conception of pain (Aydede, 2006; Tye, 2006), but are they really? Philosophers adopting the standard view have been exemplary in providing several thought experiments to putatively show the intuitive correctness of the subjectivity of pain. For example, Aydede (2006) claimed that a person who has taken an effective pain killer and claims not to feel any pain will no longer have pain. A second example was provided by Hill (2009), who stated that a wounded soldier in a battle who professed not to feel any pain did not have pain.

Reuter and Sytsma (2020) empirically investigated the subjectivity of pain by presenting participants with thought experiments. The following is the wounded-solder vignette they presented to participants.

---

While these properties are not universally accepted as indicators of phenomenality, they are at least widely endorsed. Philosophers of pain deserve some credit for identifying these characteristics in the case of feelings of pain. A third characteristic that is often discussed is that conscious experiences are private (Aydede, 2006). Reuter (2017), who drew on developmental research, argued against the privacy of pain, and Sytsma (2010) presented experimental evidence refuting the theory that pain is private.

For example, a chocolate bar can appear to be different from the way it is in reality. Thus, chocolate bars are not phenomenal states. By contrast, my visual, olfactory, and gustatory experiences of a chocolate bar cannot appear different from the way they are. Thus, these experiences are phenomenal states.
Soldiers often sustain serious injuries but show no sign of feeling any pain, continuing to display normal behavior. They also deny that they feel any pain. Only after the end of the battle do soldiers proclaim they feel severe pain and show clear pain behavior. Which of the following two statements do you consider to be correct?

(A) During the battle, the injured soldier has pain but does not feel it.
(B) During the battle, the injured soldier has no pain.

Instead of people favoring the (B) response, Reuter & Sytsma noted that 70.2% of the participants preferred (A). In almost all of their studies, the authors found that around two-thirds of the population disagreed with the standard view, believing instead that pain was not subjective. Similar results were found for the possibility of hallucinating pain. Reuter, Philips & Sytsma (2014) asked their participants to read the following prompt: “After taking the antidepressant this morning, Jenny is walking down the street when all of a sudden it feels like there is a pain in her ankle.” In response to the question “Is it possible that Jenny merely hallucinated the pain?”, 65% of the participants responded positively. Based on these and other results, Reuter & Sytsma (2017) concluded that the majority of people were likely to entertain a bodily conception of pain, according to which pain is not a conscious mental state but is actually located in the body parts. Objections have been raised by Borg et al. (2020), Salomons et al. (2021), and Liu (2021). For example, the experimental studies by Salomons et al. (2021) suggested that people’s views on feelings of pain were strongly dependent on context. Nonetheless, their results provided further evidence that the standard view is not supported by experimental data concerning the folk concept of pain, but rather that the folk concept may be polyedalic (Borg et al., 2020, see also Coninx et al., 2022) or polysemous (Liu, 2021).

Empirical research on feelings of pain is not restricted to vignette studies, nor has it been confined to studying English terms for pain. For example, Reuter (2011) conducted a corpus analysis of pain terms, and hypothesized that people distinguished the appearance of pain (that is, the feeling of pain) from the reality of pain (the pain itself) in everyday conversations. The results of his analysis revealed that, when a pain has a low intensity (for example, a mild pain), people mainly used appearance language (such as “I feel a mild pain in my arm”) but, when the pain was intense, people stated that they had (strong) pain. Such an effect of intensity matches how people make appearance-reality distinctions in other sensory domains, such as color.
vision and olfactory perception. Sytsma & Reuter (2017) replicated the corpus study of the appearance-reality distinction in the German language. The question of whether experimental-philosophical results pertaining to feelings of pain can easily be applied to other languages has received mixed results. Whereas Kim et al. (2016) compared responses from English-speaking as well as Korean-speaking populations and found no major differences, Liu & Klein (2020) argued that the pain-in-mouth argument (see Noordhof, 2001, and the experimental results by Reuter, Sienhold, & Sytsma (2019)) cannot be translated into Mandarin (but see Nie, 2021, for a different opinion regarding this problem).

3.2 Color Experiences

Although they belong to the same set of phenomenally conscious states, feelings of pain differ from experiences of color in important ways. Arguably, the three most salient differences are: (a) While feelings of pain are unpleasant, color experiences are hedonically neutral. (b) Whereas feelings of pain have a bodily aspect, color experiences do not, but seem to be located outside of the body instead. (c) Whereas pain is contingently private, colors appear to be (at least partially) publicly accessible. These differences are not minor, and are likely to influence our thinking about these conscious states beyond the obvious (see the attributions of these states to robots in Section 4.1).

In contrast to pain experts, color experts are somewhat divided regarding how to define colors. The main dividing line in philosophical research on colors is between realists, who consider colors to be mind-independent properties of objects, and relationalists, who believe colors to be mind-dependent properties (a popular version of this account states that colors are dispositions to produce certain color experiences; see Levin, 2000, and Cohen, 2009). Realists claim that introspective intuitions about how colors appear to us favor the realist account of colors (McGinn, 1996): Colors simply do not seem to be dispositional, mind-dependent properties, but rather the intrinsic properties of objects (this introspective intuition might be less persuasive when considering objects that are less mundane than tomatoes; see Adams & Hansen, 2021).

Cohen & Nichols (2010) highlighted the empirical falsifiability of claims regarding folk intuitions about how colors seem to us, and presented the results of a study that suggested that color realism might not be the dominant view among laypeople. The participants were presented with the following
vignette:

Andrew the alien and Harry the human view a ripe tomato in good light, at a distance of 1 metre. Harry says that the ripe tomato is red, while Andrew says that the very same ripe tomato is not red (in fact, he says it is green). Which of the following do you think best characterizes their views?

(1) The tomato is red, so Harry is right and Andrew is wrong.
(2) The tomato is not red, so Andrew is right and Harry is wrong.
(3) There is no fact of the matter about claims like ‘the tomato is red’.

Similar vignettes were used to test people’s views on the shape of a CD and the gustatory properties of a food item. Whereas a clear majority (69.1%) gave realist responses in the shape condition (response 1 or 2) and a clear minority (27.5%) gave realist responses in the taste condition, people were divided in the color condition (47% versus 53%). Cohen & Nichols’s interpretation of these results was that folk intuitions about colors do not clearly support the realist view.

Roberts, Andow & Schmidtke (2014), as well as Roberts & Schmidtke (2016), have challenged Cohen & Nichols’s study (mainly on methodological grounds), and have conducted their own empirical studies. Pace Cohen & Nichols, their results indicated that a majority of people (72.3% of the participants in their first study) had realist intuitions about colors. Adams & Hansen (2021) took a more cautious view of these results and criticized specific methodological aspects, but also advocated for widening the range of examples. According to Adams & Hansen (2021), a narrow focus on tomatoes and other dry, middle-sized goods was likely to provide a distorted picture of people’s color intuitions. Instead, we should also investigate people’s intuitions about the colors of rainbows, mirrors, objects in unusual lighting conditions, and so forth.¹⁰

The highly interwoven fields of the experimental philosophy of consciousness and the experimental philosophy of language also come to the fore in empirical research on people’s understanding of vision beyond the study of colors.

¹⁰The importance of the folk concept of color in philosophical theorizing about colors and color experiences has not been restricted solely to the debate between realists and relationalists. Johnston (1992) posited the existence of core beliefs that laypeople had about colors. Among others, Johnston stated that it was part of the folk concept of color that the colors of surfaces sometimes causally explain our visual experiences of things. Roberts & Schmidtke (2019) tested people’s intuitions about these core beliefs, and found that people were not in strong agreement with the latter claim.
color experiences. Although the term “see” is one of the most frequently used verbs in English, it is also highly ambiguous. Not only can “seeing” be given an informational reading as well as an introspective reading, it also has an epistemic reading. People not only see trains, they also see differences, problems, results, and truths. In a series of papers (e.g., Fischer & Engelhardt, 2017), researchers have shown that the dominant visual readings can override the contextually appropriate epistemic sense, which they have called a linguistic salience bias. Similarly, Fischer and Sytsma (2021) showed how so-called Zombie intuitions are triggered based on a problematic dominant sense of the term “Zombie”.

3.3 Emotions

In contrast to colors and pain, both experts and laypeople consider emotions to be mental states, but this does not mean that there is no controversy regarding whether emotions are phenomenally conscious states. Instead, philosophical theories of emotions have defined emotions as (a) essentially evaluative beliefs; for example, to be sad is to appraise something as a loss, (b) essentially bodily feelings; that is, to be sad is to feel that the body is drained out of energy, (c) essentially perceptions of value; for example, to be sad is to feel a loss, or (d) an amalgam of these aspects (component and composite theories).

Prior to the last five to 10 years, most of the experimental work on people’s understanding of emotions has been conducted by scientists other than philosophers. Studies have investigated the importance of facial expressions for the attribution of emotions (e.g., Wilson-Mendenhall et al., 2011), the perceived bodily locations of emotions (e.g., Nummenmaa, 2009), and the typicality and saliency of various features of emotion in self-ascriptions of emotions (e.g., Scherer & Summerfield, 1983; Fehr & Russell, 1984). In an important study by Panksepp (2000), the participants rated the feeling aspect of emotions as being more important than features such as cognitive evaluations. However, the sample was not representative of the wider population, and differences from other aspects were somewhat minor.

Given philosophers’ continued interest in determining which features of emotions are more central than others, experimental philosophers have begun to conduct their own studies that are more suited to engaging with the philosophical literature. In a recent article by Díaz (2022), the participants were asked to rate the extent to which an emotion was present after read-
ing vignettes in which individual components such as cognitive evaluations, bodily changes, and action tendencies were individually manipulated. The results indicated that none of the three features was considered to be necessary or sufficient for the attribution of an emotion, thus providing support for a prototype view of emotion concepts. However, bodily changes and action tendencies were found to depend on cognitive evaluations, suggesting that emotion concepts also have a theory-like structure.\footnote{In a corpus study investigating the use of emotion terms by both experts and laypeople, Díaz & Reuter \cite{md} found that people did in fact differentiate between being angry/happy/sad and feeling angry/happy/sad. Similar to experimental work on pain, this study suggested that laypeople made more fine-grained distinctions than the experts had expected.}

Other recent experimental studies have focused on specific emotions. For example, Kneer & Haybron \cite{2020} conducted a series of studies investigating ascriptions of happiness. In one experiment, they presented the participants with the following vignette:

George is generally very cheerful and relaxed. He is highly satisfied with his life and feels deeply fulfilled. He enjoys his life greatly and has a very pleasant experience on the whole. What George does not realize is that his wife, children and friends can’t stand him, and ridicule him behind his back. They pretend to love him only because he is wealthy. If he knew these things, he would be devastated. But they all make sure that George does not become aware of it, and in fact, he never finds out.

Question: To what extent do you agree or disagree with the following statement: “George is happy.”

Kneer & Haybron found that a majority of participants (71\%) continued to ascribe high levels of happiness to George, indicating that happiness largely tracks internal psychological factors.\footnote{The results appeared vastly different when the participants responded to a statement such as “George is doing well.”} Relatedly, a major debate in the philosophy of happiness revolves around the question of which of two psychological factors makes a person’s life happy: being satisfied with one’s life or having mostly pleasant experiences. Reuter, Messerli, & Barlassina \cite{2022} investigated whether the folk concept of happiness was aligned more strongly with affect-based states or with life satisfaction. After being presented with vignettes that manipulated the level of (dis-)pleasure and life satisfaction, a
clear majority of participants considered a person to be happy if, and only if, that person was feeling pleasurable states most of time.\textsuperscript{13}

The folk concept of happiness has also received specific attention from experimental philosophers investigating the impact of morality on attributions of happiness (e.g., Phillips, De Freitas, Mott, Gruber, & Knobe, 2017; Phillips, Misenheimer, & Knobe, 2011). The results of these studies have suggested that people’s concept of happiness is not purely descriptive, but is partly normative. The participants judged the extent to which two agents were happy to be strongly dependent on the moral lives of those agents. Díaz & Reuter (2021) extended this research to include emotions such as anger and sadness. While agreeing that the folk concept of happiness was partly normative, they believed the normativity of emotions to pertain less to moral norms and more to whether people’s emotions fit the situation in which they find themselves.

3.4 Summary

The discussed studies of feelings of pain, experiences of color, and emotions engage with fascinating but also widely different philosophical questions. Let me therefore highlight some of the main takeaways for the study of people’s intuitions and understanding of phenomenal consciousness:

- Some of the fault lines in the philosophical literature are reflected in laypeople’s concepts of conscious states: pain as a mental state versus pain as a bodily state, colors as mind-independent versus colors as mind-dependent, and emotions as beliefs versus emotions as feelings. This indicates that our philosophical theories are strongly guided by our folk conceptions. Consequently, scientific progress on our intuitions of conscious states is likely to also be able to improve our philosophical theories. Furthermore, with a more detailed picture of our folk understanding, we can not only determine more accurately where our thinking is precise and clear, but also where we are subject to biases and misperceptions.

- Experimental research has shown that people differentiate between pain and feelings of pain, colors and color experiences, and emotions and

\textsuperscript{13}Other emotional states that have recently been investigated empirically include the state of being moved (Cova & Deonna, 2014), as well as guilty pleasures (Goffin & Cova, 2019).
emotional feelings. These results suggest that laypeople seem to focus more on bodily states (in the case of pain), objects (in the case of colors), and cognitions (in the case of emotions). Consequently, the frequency with which people think about phenomenal consciousness is likely to be much lower than hitherto assumed. This also has important consequences for the scientific study of conscious states: Researchers need to be more attentive to the way in which questions are presented to participants.

- Although our folk concepts of conscious states provide some guidance for philosophical theories, some of the empirical results have produced some surprising facts. For example, (i) laypeople make more fine-grained distinctions than philosophers had presumed (Reuter & Sytsma, 2020), (ii) introspective evidence of conscious experiences may not be a good guide to the concepts of consciousness that we entertain (Cohen & Nichols, 2010), and (iii) moral considerations influence our thinking about phenomenal consciousness (Phillips et al., 2017). These and other results are likely to occupy philosophers for some time to come.

- The experimental results that were reviewed demonstrated tremendous intra- and interpersonal variations in the ways in which laypeople think about phenomenal consciousness. Thus far, we do not know which factors drive these differences. How context dependent are people’s responses? How does knowledge about conscious states influence our intuitions about consciousness? Does the way a conscious state feels to a person, such as some people experiencing happiness more intensely than others, influence their understanding of that conscious state? Does the frequency with which conscious states are felt have any impact? For example, do patients suffering from chronic pain have the same concept of pain as healthier people? Answering these questions requires a massive collaborative effort of linguists, philosophers, psychologists, and other cognitive scientists.

This short summary concludes the first section, in which I reviewed some of the experimental-philosophical work on individual conscious states. In the following section, I will consider studies that pertain to illuminating people’s views and understanding of phenomenal consciousness more generally.
4 The Folk Concept of Consciousness and Attributions of Consciousness

Many people who are familiar with the experimental-philosophical literature will associate the phrase “experimental philosophy of consciousness” most strongly with the following two research questions: (a) Which entities, such as snails or robots, do laypeople consider to be (phenomenally) conscious? (b) Is there an implicit folk concept of consciousness? Empirical research has shown that these two questions are closely linked, thus leading to a flourishing research program concerning both of these questions. Before I present the results of three central papers (Gray, Gray & Wegner, 2007, Knobe & Prinz, 2008, Sytsma & Machery, 2010) that have triggered much of the subsequent research on the folk concept of consciousness and attributions of consciousness, let us quickly recap the intentional-phenomenal framework that provides the background for both questions.

Philosophers commonly identify two distinct marks of the mental, namely intentionality and phenomenality. A mental state that is intentional is directed toward a real or imaginary state of affairs, while a mental state that is phenomenal feels like “something to be in”. Some states have been argued to be exclusively phenomenal (such as feelings of pain and moods), some states to be exclusively intentional (such as beliefs and mental calculations), and others to be both phenomenal and intentional (for example, desires and sensory experiences).

A natural question that arises for the experimental philosopher is: Do laypeople differentiate between phenomenal and intentional states? If they do not, this would raise some serious concerns about whether laypeople actually possess the concept of phenomenal consciousness. Remember that the problem of the missing folk term rules out an explicit folk concept of consciousness. It follows that, if laypeople were to intuitively group mental states according to those that have phenomenality and those that do not, this would at least support the theory that there is an implicit folk concept of consciousness. If they do not, philosophers have some explaining to do: How can it be that phenomenal consciousness is “the most central and manifest aspect of our mental lives” (Chalmers, 1995, p. 207) if no one except philosophers has really noticed it?

The extensional bottom-up strategy that was introduced above allows researchers to tap into people’s implicit concepts about the mental realm.
The basic assumption underlying this strategy is as follows: If an entity like a sophisticated robot or a turtle is conceived of as having one type of phenomenally conscious state, such as feeling pain, people are likely to also conceive of the entity as having other phenomenally conscious states, such as experiencing hunger or experiencing red. Thus, if laypeople attribute an entire range of phenomenally conscious states to one group of entities but deny these states of consciousness to another, this would provide quite strong evidence that laypeople have the same implicit concept of consciousness that philosophers explicitly entertain.

4.1 The Three Initial Papers

Given these preliminaries, the stage is set for some of our main actors to enter the scene. In 2007, Gray et al. published a short paper in *Science* that argued that laypeople seemed to divide mental states into two groups, which the authors called EXPERIENCE and AGENCY. In their study, over 1200 unique participants were asked to state which of two entities was more likely to be able to have a certain mental state; for example: “Which is more capable of feeling hunger: a frog or a fetus?” The entities under investigation were humans in different developmental stages (fetus, baby, adult and a dead person), a robot, god, and some animals. The mental states about which people made judgments included prototypical phenomenal states (feeling pain, fear, hunger, and rage) and prototypical states and processes involving intentionality (thought, planning, recognition, and memory). The authors found that entities that were considered to be capable of feeling hunger and fear were also considered to be capable of feeling pain and pleasure. The outcome of this study provided support for thinking that laypeople do indeed make the same two-fold distinction (intentionality versus phenomenality) that is central to the philosophy of mind.

Prinz & Knobe’s (2008) paper on group consciousness also concluded that laypeople divided the mental realm into states harboring phenomenal consciousness and states that were primarily intentional. More specifically, Prinz and Knobe’s studies suggested that laypeople were willing to ascribe intentional states to fictional (for example, Acme Corporation) as well as real (such as Microsoft) group agents, but refrained from attributing phenomenal states

---

14 Both the factors EXPERIENCE and AGENCY accounted for a total of 97% of the variance in the data.
to such group agents. Corpus data concerning the frequency of phrases such as “Microsoft intends/decides” versus “Microsoft feels happy/pain” revealed that only the former phrases were used frequently, whereas phenomenal attributions did not occur. Additional support for this division was found in a vignette study, in which the participants considered sentences ascribing intentional states to be far more natural. For example, the statement “Acme Corporation intends to release a new product this January” received a mean rating of $M=6.3$ on a scale from 1 (“sounds weird”) to 7 (“sounds natural”) compared to sentences attributing phenomenal states (“Acme Corporation experiences great joy”, $M=3.7$). As the results could have been driven either by terms indicating phenomenal consciousness, such as “feeling”, or by terms referring to states associated with emotions, the researchers conducted a further study to test for both factors. The participants clearly considered statements such as “Acme Corporation regrets its recent decision” to be more natural ($M=6.1$) than “Acme Corporation feels regret” ($M=2.8$). This outcome not only provides some evidence that phenomenally conscious states are not attributed to group agents, but also that laypeople appear to make an implicit distinction between phenomenal and intentional states.

The third of the three initial papers dealt the first serious blow to the idea of an implicit folk concept of phenomenal consciousness. Sytsma and Machery (2010) used the same extensional bottom-up approach, but included an important item that was absent in the other two papers: seeing colors. Sytsma and Machery hypothesized that, if the folk concept of phenomenal consciousness matched the philosophers’ concept, we should find that people believe that humans see colors and feel pain, but would not attribute these states to a robot. However, this was not what they found. In the robot conditions, the participants read a vignette in which a robot either discriminated colors or avoided pain stimuli. Crucially, laypeople answered that the robot saw red ($M=5.15$ on a scale from 1 (“clearly no”) to 7 (“clearly yes”)), but denied that the robot felt pain ($M=2.54$). By contrast, participants with a philosophical background disagreed with laypeople and mainly denied that the robot saw red ($M=3.48$). The latter result, according to Sytsma & Machery, might well indicate that a philosophy education changes folk intuitions about consciousness. The authors also proposed their own positive account, the valence account, according to which laypeople’s concept

\[ ^{15} \text{Using a similar experimental paradigm, Huebner et al. (2010) replicated the results obtained by Knobe & Prinz, but also found significant cross-cultural differences.} \]
of consciousness only refers to states that feel pleasant or unpleasant to an agent\textsuperscript{16}

4.2 Criticism and Subsequent Research

More recent studies investigating how many and which dimensions of mind perception there are have increased the pressure on Gray et al.’s two-dimensional picture. While Weisman et al. (2017) appreciated Gray et al.’s bottom-up approach, they were critical of the fact that the participants only answered questions about a single type of mental state for a series of different entities. For example, the participants first compared a frog and a god in terms of experiencing pain, then compared a god and a dead person in terms experiencing pain, and so on. In order to reveal a more precise picture of mind perception, Weisman et al. had their participants only consider a single entity, such as a beetle or a robot, but then had them rate the ability of that entity to experience a wide range of mental states such as experiencing joy, having desires, and making choices. The results did not reveal a two-dimensional picture, but suggested that laypeople classified mental states according to three categories, which the authors called BODY, HEART, and MIND. The factor BODY was dominant for mental states that were related to bodily aspects such as feeling hungry, feeling pain, and experiencing fear. Mental states that had a social component such as feeling embarrassed, proud, and disrespected were strongly grouped together by HEART. The factor MIND corresponded to perceptual-cognitive abilities such as seeing and remembering things, making choices, and the like. Of note, while some similarities between the two- and three-dimensional frameworks emerged, the experiential and agentic capacities in Gray et al. were distributed among Weisman’s three categories. Malle (2019) further simplified the questions posed to participants: All mental states and processes $X$ had to fit the question schema “Is the agent capable of $X$?”, and expanded the item pool substantially. The results revealed the replication of a three-dimensional structure that was not widely different from Weisman et al.’s study: Malle identified three dimensions that were roughly equivalent to Weisman’s BODY, HEART, and MIND dimensions.

\textsuperscript{16}The results of an additional experiment suggested that valence of smell was a driving factor for attributions of consciousness: Robots are considered capable of experiencing odorless smells, but less so vomit or bananas. Following further experiments, Sytsma (2012) abandoned the valence account.
Similar to Gray et al.’s research, Knobe & Prinz’s results were subject to substantive criticism. The arguably most problematic aspect of Knobe & Prinz’s research was an imbalance in the stimuli, as they did not control for the effect of specifying the intentional context. Arico (2010) designed an experimental study in which the contributions of the intentional context of a statement, as well as the contributions of the inclusion of the term “feeling”, could be measured more reliably. To accomplish this, he presented the participants with statements such as “McDonald’s / Donald is feeling upset” (feeling, no intentional context), “McDonald’s / Donald is upset” (no feeling, no intentional context), “McDonald’s is feeling upset about the court’s ruling” (feeling, intentional context), and “McDonald’s is upset about the court’s ruling” (no feeling, intentional context). Pace Knobe & Prinz, the results revealed that the inclusion of the term “feeling” did not have any effect on how strange or natural an attribution of an emotional state sounded to the participants. By contrast, providing an intentional context, such as “about the court’s ruling”, had a substantial effect on people’s ratings. In line with Knobe & Prinz, Arico did find an effect of agent (human versus group agent), although this was primarily driven by the “without context” conditions; that is, the participants found it much stranger to say that McDonald’s was upset than it was to say that Donald was upset. Other aspects of Knobe & Prinz’s approach were also critiqued. Phelan, Arico & Nichols (2013) questioned whether people do in fact attribute mental states to group agents, or whether they attributed them distributively to members of the group. Using a pronoun replacement task, Phelan et al. (2013) found that people were more likely to select the plural pronoun for mental states, but found no difference between intentional and phenomenal states.

There was also no shortage of criticism of and responses to Sytsma and Machery’s research. Most of the objections focused on the wording of the crucial item in Sytsma and Machery’s main study, in which the authors asked their participants the target question “Did Jimmy see red?” Several authors have noted that the phrase “seeing red” is ambiguous or at least polysemous as it allows two different readings (e.g., Peressini, 2014; McLaughlin & Rose, 2018; Chalmers, 2018). In the phenomenal sense, to see something that is red entails having a visual experience of red. However, in the informational sense, seeing red only entails being able to discriminate red from other colors based on visual information collected from the environment. It is plausible to assume that robots can see red in the informational sense but not in the phenomenal sense. If the participants interpreted the wording of the
target question in the informational sense, this could easily explain the high ratings when attributing seeing colors to robots. To address this objection, Sytsma and Ozdemir (2019) asked participants “Did Jimmy experience red?” The average rating was not significantly different from the rating for the original phrasing of the question. While these results do suggest that people interpret “seeing red” in a phenomenal sense, it might still be the case that the participants answered differently in cases in which colors were dreamed about or hallucinated; that is, situations in which their focus was directed inwards.

4.3 Mind-Body Dualism, the Hard and the Meta-Problem of Consciousness

The question of whether or not the mind is distinct from the body dwarfs all other questions in the philosophy of mind. Although materialism is now the most widely accepted position, many philosophers (even many materialists) still insist that we face the hard problem of consciousness; that is, the problem of explaining how states of the brain give rise to phenomenal consciousness (e.g., Chalmers, 1995). At the heart of the hard problem of consciousness are a range of dualist intuitions that are fueled by thought experiments, such as Mary’s room and the Zombie case. Thus far, it has not been established whether dualist intuitions are shared by laypeople with no prior exposure to philosophical arguments. Such insights may not only help us to understand why the hard problem persists (Papineau, 2020), but are also part of revealing people’s beliefs about consciousness. Note that we are not interested in asking whether people conceive of the mind or the soul as being distinct from the body, as this question is likely to be answered in the affirmative by those who believe in an afterlife (e.g., Bloom, 2005). However, it would be a mistake to think that, just because people believe in a soul, they also believe that phenomenal consciousness is part of that soul. Phenomenal consciousness might be thought to be purely material, and the soul may merely be able to enjoy these phenomenal-cum-physical states.

Most of the existing research on mind-body intuitions has examined whether children start out with monistic or dualistic tendencies. The available studies suggest that children are intuitive dualists. For example, Hood et al. (2012) asked children aged six to eight whether a duplicated hamster (they were presented with a flashing machine that “copied” soft toys and
hamsters) would know the name of the child. While all the children thought that the original hamster did know their names, around half of the children did not ascribe this knowledge to the duplicated hamster. Forstmann & Burgmer (2015) adapted Hood et al.’s design to test adults’ intuitions regarding physical properties such as having a scar, intentional states such as knowing when it is time to eat and, most importantly for our purposes, phenomenal states such as being afraid of the laboratory intern Mike. The participants were told that the duplicated hamster was a 100% identical copy of the original hamster, and were asked to state the degree to which each attribute applied to the original and to the duplicate hamster using a seven-point scale. Forstmann & Burgmer found a drop of $\Delta = 1.15$ in the mean ratings for the mental attributes (they excluded participants who also assigned lower ratings for the physical attributes), and concluded that adults intuitively dissociated minds from bodies. One might wonder whether a drop of $\Delta = 1.15$ supports such a strong conclusion, particularly when taking the fact that the mean rating was still significantly above the midpoint into account. Furthermore, intentional and phenomenal states were grouped together in their analysis.\[17\]

Sytsma and Snater (2021) conducted a study that was similar to the one by Forstmann & Burgmer. Although their focus was on how different target mental items clustered together, their study revealed interesting insights into whether people are phenomenal dualists or phenomenal monists. The participants were given the following prompt: “Imagine that scientists scan you and use that information to create an exact physical duplicate of your body (including your brain). What, if anything, do you think this duplicate would be capable of?” The results provided quite strong support for intuitive dualism. For example, the mean rating for “The duplicate would feel pain when she is injured” was 4.25, and “The duplicate would have dreams when she sleeps” was rated 3.98 on a seven-point scale. However, as Sytsma and Snater did not test for physical attributes, we lack an important control condition.

Apart from the hard problem of consciousness, the past few years have seen an increase in discussions of the meta-problem of consciousness, which is the problem of explaining why “we think consciousness poses a hard prob-

\[17\]Forstmann & Burgmer (2022) also conducted an experimental study testing laypeople’s views on the location and distribution of consciousness in the brain using a drawing task, and found that people mainly located consciousness in a confined area of the prefrontal cortex.
lem” (Chalmers, 2018, p. 6). Sytsma & Ozdemir (2019) extended their skepticism about the existence of a folk concept of phenomenal consciousness to also support skepticism about the meta-problem of consciousness. Chalmers rejected this conclusion by highlighting that the meta-problem of consciousness also arises for individual conscious states such as feeling pain. Díaz (2021) adopted a more direct approach for testing the meta-problem of consciousness by asking participants questions such as “The properties of pain are fully explained in terms of neural activity in the DPI”. He found that a majority of 81% of the participants considered neural activity to fully explain feelings of pain, thus raising doubts about whether the meta-problem is a problem for laypeople.

4.4 Summary

4.4.1 The Folk Concept of Consciousness

The problem of the missing folk term makes it necessary to investigate people’s implicit understanding of consciousness. Researchers have used the extensional bottom-up strategy to determine how many and which dimensions of mind are perceived by laypeople. Initial results by Gray et al. (2007) and Knobe & Prinz (2008) provided promising support for a two-dimensional picture that matched the philosophers’ distinction between intentionality and phenomenality. More recent results have painted a different picture that seriously questions whether laypeople have an implicit concept of phenomenal consciousness. Not only do laypeople make a three-fold distinction of the mind, they also fail to group seemingly prototypical phenomenal states.

However, there are reasons to be less “pessimistic” about the folk conception of consciousness. Philosophers have acknowledged that some mental states are likely to have both intentionality and phenomenality. As shown in Section 3.3., emotions might well be paradigm cases of such composite states; note also that emotions appear to be dominant in the third dimension of the mind (for example, Weisman et al.’s HEART dimension). In other words, a three-dimensional picture might well be exactly what we should expect given the two markers of the mental state (also see Wyrwa, 2022).

Perhaps more concerning is the observation that people group experiences of colors and sounds with intentional states and not with other phenomenal states. However, research by Malle (2019), Ozdemir (2021), and Sytsma & Snater (2021) has shown that, with regard to dreaming and imagining
things vividly, people group these states more closely with other phenomenal states. This suggests that people may well consider the experience of colors and sounds to be phenomenal states similar to the feeling of pain, but that such a classification only comes to the fore when people focus on the internal aspects (colorful dreams) and less on the external world (colorful objects).

4.4.2 Attributions of Consciousness

Mind perception studies allow us to draw conclusions regarding which entities people believe to have phenomenally conscious states. Babies, adults, and mammals are clearly considered to have a wide variety of phenomenally conscious states. Fetuses, PVS patients, beetles, and frogs receive lower ratings in comparative studies, but are still considered to experience at least some phenomenally conscious states (such as feeling pain) to a certain degree. God, dead people, silicon robots, artifacts, and group agents such as Microsoft are generally not considered to be capable of experiencing phenomenally conscious states. The case of robots is more difficult to assess given the results for color and sound experiences.

Knobe & Prinz (2008) explored two possible factors that might influence people’s attributions of consciousness. First, people attribute a phenomenally conscious state if the entity realizes the functional role (that is, the causes and effects) of that state. Second, the physical constitution of an entity influences people’s attributions of consciousness: If an entity has the correct biological make-up, people consider the entity to be phenomenally conscious; otherwise, they do not. Knobe & Prinz interpreted their studies of group agents (similar functional roles but no biological make-up) to show that physical constitution played a much more important role in the attribution of phenomenal states than had previously been acknowledged. While many of the claims that were discussed in this paper have been very controversially received, there seems to be some consensus in the literature that physical constitution is a central factor in the attribution of conscious states (but see Phelan & Buckwalter, 2012, for a different view).

5 Concluding Remarks

I began this article by highlighting that we do not have a term for consciousness that is part of our folk terminology; in other words, the problem
of a missing folk term. Given this problem, as well as the widely differing properties that different conscious states have, our understanding of people’s thinking about the phenomenal mind proceeds in piecemeal fashion. Results from studies of people’s intuitions about feelings of pain, color, and sound experiences, as well as feelings of anger and happiness, need to be compiled and interpreted. No sweeping and grandiose claims are likely to appear from a single study.

However, these problems also present tremendous opportunities. First, experimental philosophy of consciousness has become a methodological hotspot. Knobe & Prinz (2008) and Reuter (2011) were among the first to use corpus analyses for philosophical investigations. Fischer & Engelhardt (2017) have broken new ground by using pupillometry and eye tracking to investigate people’s understanding of perceptual terms, and the extensional bottom-up strategy that was discussed at length in this paper continues to deliver illuminating insights into people’s understanding of consciousness.

These problems also make us think more deeply about the relationship between folk theories and philosophical theories. While laypeople might be massively mistaken or even ignorant about phenomenal consciousness, we cannot develop promising theories about feelings of pain, color experiences, and emotions without detailing the content and structure of the corresponding folk concepts. How folk intuitions at the basic level may shape our theories at the superordinate and superduperordinate levels is far from clear at this stage.

Lastly, the challenges we face when studying people’s understanding of phenomenal consciousness should make the philosophers’ hearts beat faster. Contradictory intuitions need to be addressed, ambiguous language needs to be clarified, and messy empirical data need to be squared with clean philosophical theories. If this is not a reason to be excited about what is to come in the next 10 to 15 years, I do not know what is.

References


