Measuring the mental experience

Michael Pauen\textsuperscript{a,}\textsuperscript{*}, John-Dylan Haynes\textsuperscript{b,c}

\textsuperscript{a} Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Department of Philosophy, Unter den Linden 6, 10099 Berlin, Germany
\textsuperscript{b} Charité-Universitätsmedizin Berlin, Bernstein Center for Computational Neuroscience, Charitéplatz 1, 10117 Berlin, Germany
\textsuperscript{c} Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Unter den Linden 6, 10099 Berlin, Germany

\textbf{A R T I C L E I N F O}

\textbf{Keywords:}
Consciousness
Third-person access
Introspection
Explanatory gap
Measurement
History of science
Pain
Metacognition, Binocular rivalry

\textbf{A B S T R A C T}

Many philosophers have argued that the subjective character of conscious experience results in a fundamental deficit of third-person (henceforth: extrospective) access to first-person experience. By comparing extrospective measurement techniques with measurement techniques in the natural sciences, we will argue that extrospective methods suffer from no such deficit.

After a rejection of some principled objections against extrospective methods, a historical comparison with the development of measurement techniques in the natural sciences will show that extrospective measuring methods are still in an early stage of development. However, they can be significantly improved by way of a bootstrapping strategy, similar to that which has proven successful in the development of physical measurement techniques. One reason to expect such improvement is the availability of multiple sources of evidence, which should allow for substantial advances in extrospective measurement techniques. Finally, we will discuss new developments in pain measurement in order to show that the bootstrapping strategy is already bearing fruit.

\section{Introduction}

Mental experiences appear to be essentially subjective. I experience my pain in a way that no one else does. Apparently, I have an \textit{experiential} privilege regarding my own mental states. Many philosophers and scientists conclude that the essential subjectivity of the mental deeply affects our epistemic access to it: In their view, I am in a much better position than anybody else to acquire \textit{knowledge} regarding my own experience. Conversely, objective scientific methods seem to suffer from a fundamental epistemic deficit regarding the mental, particularly because they lack direct access to their object.

The problem found one of its most succinct analyses some decades ago in Thomas Nagel’s work (Nagel 1974, 1986). According to Nagel, we find ourselves confronted with a dilemma regarding the mental. If we want to meet objective scientific standards, then we will never be able to capture the subjective character of mental experience. Conversely, if we want to do justice to the subjectivity of the mental, then objective scientific standards seem out of reach. Nagel thinks that scientific access to conscious mental states is possible, but it misses out on what is essential for consciousness: subjective experience.

While this diagnosis reflects an intuition that goes back to ancient philosophy (Cary 2000), it has consequences that continue to affect today’s research, as objective measurement methods gain importance in studies on metacognition, other higher cognitive states, and theories of consciousness. This can be seen in the debates about \textit{Cognitive and Non-cognitive Theories of Consciousness} (Overgaard

\textsuperscript{*} Corresponding author.
\textit{E-mail addresses:} m@pauen.com (M. Pauen), haynes@bccn-berlin.de (J.-D. Haynes).

https://doi.org/10.1016/j.concog.2021.103106

Received 18 June 2020; Received in revised form 18 February 2021; Accepted 19 February 2021
Available online 16 March 2021

1053-8100/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
and Grünbaum 2012, Overgaard 2015), the Neural Correlates of Consciousness (Chalmers 1998, Metzinger 2000, Overgaard and Overgaard 2010, Koch et al. 2016) as well as in methodological analyses like the one on No-report Paradigms – particularly in metacognition studies (Frässle et al. 2014, Michel 2017, Block 2019), but also in discussions about first-person experience and first-person reports in neuroscience and psychology (Jack and Roepstorff 2002, 2003, Schwitzgebel 2002a, b, 2008). Potential limitations of third-person methods become even more relevant insofar as these methods are put to work in clinical contexts, e.g., in order to detect phenomenal experience in behaviorally non-responsive patients (Naci, Sinai, and Owen 2017) or to assess pain intensity with the help of biomarkers (Wager et al. 2013, Woo, Chang, et al. 2017). Finally, the assumption of a fundamental epistemic deficit of third-person methods seems to support the idea of an Explanatory Gap, which, if sustained, will preclude us from ever providing a scientific theory of conscious experience (Levine 1983, Chalmers 1996).

Here we will argue that third-person methods suffer from no such fundamental epistemic deficit regarding the mental. Actual epistemic deficits exist at present, but they can be explained by contingent factors, such as the stage of development of measurement techniques or the complexity of the measurands.

Similar claims have been argued for already, most notably by Dennett in his account of Heterophenomenology (Dennett 1988, 2003, 2005, see also Irvine 2019). But unlike Dennett and other previous proponents of similar views, our analysis corroborates this claim with a systematic and historical comparison between measuring procedures regarding the mental and those elsewhere in science. Moreover, against qualia skeptics like Dennett, we will demonstrate that even first-person experiences with a specific phenomenal character can be measured with valid objective methods.

In what follows, we will provide an epistemic analysis of third-person, scientific methods directed at first-person subjective experience. We will refer to these methods as “extrospective,” and our main claim is that they do not suffer from a fundamental epistemic deficit.

Our argument proceeds in four steps. In the first step (Section 3), we demonstrate that some familiar in-principle objections that allegedly demonstrate the fundamental deficit of extrospective methods should be dismissed. While this is a far cry from proving our main claim, it paves the way for a more detailed analysis, which will strongly suggest that the remaining deficits are contingent. We start motivating this analysis in step two (Section 4), where we provide a historical and systematic comparison of extrospective and physical measurement techniques. The comparison demonstrates that certain problems that are taken to be peculiar to extrospective measuring methods were, in fact, observed in the initial stages of typical physical measurement techniques as well. This indicates that extrospective methods may still be in an early stage of development, which would at least partially explain the present shortcomings of these methods. In the third step (Section 5), we further corroborate our claim that contingent factors explain these deficits, showing that significant progress can be made with respect to extrospective methods, partly because the bootstrapping strategy that has proven successful in the development of physical measurement techniques can be applied to extrospective methods as well. One important basis for the possibility of bootstrapping is the availability of multiple sources of extrospective evidence, which allows for the detection of errors even in introspective reports, and for the improvement of measuring methods. Finally, in the fourth step (Section 6), we discuss new developments in pain measurement in order to show that the strengths of extrospective methods are already increasing, and that there is no reason to think these methods are not able to capture first-person experience.

Our conclusion is that the subjectivity of conscious experience does not translate into a fundamental epistemic deficit of objective extrospective knowledge of this experience. Rather, there is a basic epistemic symmetry between physical and extrospective methods, particularly if we account for contingent factors like the stage of development of extrospective methods and the extreme complexity of the mental.

2. Conceptual clarifications

Let’s start with a clarification of the concept of extrospection, which we have already briefly characterized above as third-person epistemic access to first-person experience. As this concept is quite new, we will provide a more detailed explanation here.

(1) First, we distinguish between (a) a mental experience, say a pain state, on the one hand, and (b) our knowledge of it, including our efforts to acquire this knowledge, on the other. Having a pain experience and knowing that one has it are two different things, at least conceptually (Chalmers 2003).

Note that this conceptual distinction does not exclude that, as a matter of fact, experience and introspective knowledge about it might never come apart. If this is the case, then the conceptual distinction would be needed to express this very fact. This, however, would imply the infallibility of introspective beliefs, which is highly controversial. Claims of infallibility or quasi-infallibility, if they are made at all in current philosophy, are restricted to very specific cases (Chalmers 2003, Gertler 2012).

(2) Second, the person having or acquiring the knowledge about an experience can either be identical with the person having the pain experience or it can be someone else. In the first case, the knowledge of the experience and the related efforts count as introspective. Introspective knowledge is knowledge about one’s own experiences and it is based on one’s own cognitive abilities.

Conversely, a piece of knowledge of an experience and the efforts to acquire it count as extrospective if the person having or acquiring the knowledge is not identical with the person having the experience. In short: Extrospective knowledge is about others’ mental states. Extrospection uses methods like occasional observation, mind-reading, systematic behavioral measurements, physiological measurements, neuroscientific data, but also others’ introspective reports.

There are fringe cases when a person uses external methods, e.g., fMRI, in order to acquire information about their own experiences. While it seems clear that this does not count as introspection, we admit having no univocal intuitions whether it should count as extrospective either.
(3) Third, we need to distinguish different categories of evidence about a person’s experience, particularly if the evidence is based on introspective reports.

(c category i) Explicit Reports

Introspective reports are explicit if they directly convey the information that is needed, e.g., “I feel pain” in a study that measures whether or not subjects feel pain. Typical formats are verbal reports, pressing a button, ticking a box in a questionnaire, or marking a point on a continuous scale.

(c category ii) Implicit Reports

Introspective reports are implicit if they only indirectly convey the information that is needed, such that additional steps are required to extract information the participants are not aware of. For example, participants may explicitly report whether a stimulus has changed its intensity in a psychophysics experiment. This report is then used, in a second step, to calculate the relations between stimulus strength and felt intensity, which is the information that is desired. This requires a theory or an algorithm in virtue of which an experimenter can infer facts about experience from facts about behavior; the reports or button presses are just data points.

(c category iii) Objective Methods

A third category utilizes a host of behavioral, physiological, and neural measures instead of introspective reports. Among them are reaction times, blushing, behavioral observation, electrodermal activity, EEG, and fMRI data. The interpretation of these data requires a theory connecting the objective findings with first person experience which calls for a calibration based on introspective report.

Following a widely shared but somewhat liberal view that goes back to Helmholtz and Stevens (1946), we understand measuring as the process of assigning numbers to properties. We also follow Stevens' fourfold distinction of measuring scales: Nominal scales use numbers as mere labels to distinguish properties (e.g., male vs. female; the days of the week) without imposing any further order or quantification. Ordinal scales use numbers to impose a rank order on the expressions of a property (e.g., the Mohs scale of the hardness of minerals), but do not quantify the differences between the individual steps. Such quantification does occur in interval scales, like the Centigrade or Fahrenheit temperature scale. These scales use numbers to quantify properties such that the differences between distinct values become meaningful, but as they lack an absolute zero point, we cannot say that a fluid at 20 °C is twice as hot as a fluid at 10 °C. This is possible when we use a ratio scale like the absolute temperature scale (Kelvin) which does have an absolute zero point (Stevens 1946).

Our paper will focus on a comparison between extrospective methods for measuring mental phenomena, on the one hand, and measuring techniques in the natural sciences, on the other. We will use two criteria: Validity, which evaluates whether a given method captures the phenomenon that it is supposed to measure (Messick 1995, Irving 2018), and accuracy, which describes the proximity of a given measuring result to the true value of the magnitude in question (JCGM 2012).

We will call the epistemic deficit of a given method fundamental if it meets the following two conditions: First, the deficit is inherently tied to the method in question. For example, extrospective methods would meet this criterion if their deficits were due to the incompatibility of the method’s third-person nature with the subjective character of conscious experience. Second, the deficit makes a difference in kind, not only a difference in degree. A difference in kind between two methods, A and B, means that results from A will practically always trump results from B, and/or that basic epistemic requirements which are met by A are not met by B.

Following a familiar understanding (Bell 2010, Searle 1995, 7sq.), we will apply the objective/subjective distinction both in an ontological and an epistemological sense. Colors are thought to be ontologically subjective because subjects perceiving colors are essential for what colors are. By contrast, whatever makes up the beauty of a landscape seems to exist independently of any subject watching it. But recognizing the landscape as beautiful depends on the taste of humans making a judgment. This is why beauty is taken to be epistemically subjective (Bell 2010, Nagel 1979). Moreover, we assume that a given conscious experience can be called subjective because it is restricted to a specific subject. Frequently, however, subjective is used in the literature as a synonym for the phenomenal character of a given experience. In order to avoid confusion, we will use expressions like phenomenal or qualitative experience instead in these cases, as they are less ambiguous.

3. Principled objections: Epistemic directness and public access

For quite a while, the discussion about the epistemic merits of extrospection has been focused on principled objections against extrospection. These objections tried to show that extrospection suffers from a fundamental epistemic deficit not only in comparison to introspection, but even more so in comparison to standard third-person methods. According to the first objection, extrospection suffers from such a fundamental epistemic deficit because, unlike almost any other epistemic method, it lacks direct access to its objects. In a similar vein, the second objection states that extrospection fails an essential requirement of any scientific method, namely public access, because the mental facts it is about are only privately accessible.

After a brief survey of the history of the concept of “consciousness,” including its use in other language communities (3.1), we will advance a suggestion about how to understand directness (3.2). We will then discuss the first objection, arguing that indirectness neither sets extrospection apart from other scientific methods nor is it, taken by itself, an epistemic deficit (3.3). In the following subsection (3.4) we will discuss the second objection, showing that extrospection meets the requirement of public access.

Note that our dismissal of these objections does not show that extrospection is free from any fundamental deficit. In order to do so, a more detailed epistemological analysis is required. Such an analysis, which focuses on a comparison of extrospective measurement with standard measurement methods in the natural sciences, will constitute the main part of this paper in Sections 4-6 below.
3.1. Directness and the concept of consciousness

Directness is one of the most important issues in the debate about the epistemic merits of extrospection, particularly in comparison to introspection. Directness is thought to be essential for the specific epistemic role performed by introspection. This view can already be found in ancient philosophy, particularly in Plotinus and Augustine (Cary 2000), and plays an important role in discussions on consciousness in the 17th and early 18th century led by authors like Descartes (1984), Cudworth (1731), Locke (1836, 392), Leibniz (1890, 221), Zachary Mayne (1728) and, later, Brentano (2009).

In some way or other, these authors assume that having a conscious experience is directly related to having knowledge of this experience. Thus Mayne (1728, 175 sq.) argues that “the Mind’s Consciousness of all its Acts, … does immediately accompany, and … closely adhere to them. … Conscious Knowledge or Perception is also most perfectly and thoroughly adequate and exact.”

This intimate relation between conscious experience and knowledge of this experience—which sometimes defies the distinction between experience and knowledge made in the conceptual clarification above—is already evidenced by the very term “consciousness” and its synonyms in all the languages in which the concept became relevant in the 17th century. These terms are all compounds of the word “knowledge” and a preposition, typically “con” or “co”: Con-scientia, con-science, con-sciousness, co-scienza, con-scienza, and Be-wusstein. Originating from the Latin word “consīus” which denotes a confident or witness, “conscientia” in both its moral (consciousness) and its psychological (consciousness) sense is thought to provide substantial or even “incorruptible knowledge” (Hennig 2007): “Conscientia est … certīlīma scientia & … certitudo eius rei quae animo nostro inest: sive bonum sit, sive malum.” (Consciousness is … the most certain knowledge and … certainty of what is in our mind, be it good or bad) (Thesaurus 1578, Lamarra and Verde, 2014). While conscientia as moral conscience has played an important role already in the 16th and 17th centuries (Hennig 2007), it took the efforts of Descartes and other philosophers following him to establish the psychological concept of consciousness as well as the related idea of an introspective privilege in epistemology and the philosophy of mind.

It should be noted that the conception of the mental suggested by the concept of “consciousness” and its early proponents differs significantly from that denoted by the more ancient and encompassing notion of the “soul,” varieties of which can be found across many different Western and non-Western cultures (Sheils 1978). “Soul” refers to a substance, typically of supernatural origin, that explains a host of perceptual, emotional, cognitive, and volitional abilities, it stands for the identity of a person, but it also explains life. That is why the majority of terms referring to the soul are related to some sort of breath, among them “psyche”, “pneuma”, “atman”, “spiritus” or “fatus”, which are “expired”, once the life is over. By contrast, epistemic aspects related to reflection or privileged self-knowledge that become dominant with the concept of consciousness play a minor role, if any at all, for the idea of the soul.

Something similar can be found when we look at mentalistic concepts in non-Western languages even if they have been rendered as “consciousness.” Cross-cultural comparisons are difficult, though (Throop and Laughlin 2007), since many of these concepts have a broad spectrum of meanings and hardly any of them corresponds directly to “consciousness.” Concepts from the Indian “science of consciousness” (Kak 1997) like “mana”, “citta” and “vijñāna” that are translated as “consciousness” (Dreyfus and Thompson 2007, Bodhi 2000) differ from the English concept not only due to their close association with perceptual abilities (Bodhi 2000, 94 n 154), but also because many of them (e.g., vijñāna) are taken to denote a “life-principle,” much like our concept of the “soul.” Even more importantly, none of these notions shares the far-reaching epistemic commitments implied by the Western concept of “consciousness.” As Kathleen Wilkes (1988) has shown, the same holds for the Chinese notion of “yi-shih” or “svijest” in Croatian, even if they are translated as “consciousness” as well. Finally, the concept “noman” in the Melpa language spoken in Papua New Guinea, another counterpart of “consciousness,” encompasses notions of will, agency, intention, and sensibility (Stewart and Strathern 2000), but also lacks the epistemic commitments of its counterpart in English.

3.2. What is Directness?

Let’s conclude, then, that the English concept of consciousness and its attendant epistemic commitments are highly distinctive, compared both to related concepts in the earlier Western tradition, and to similar notions in non-Western language communities. These epistemic commitments entail a strong introspective privilege which is thought to be based on direct first-person access. Interestingly, this claim to introspective privilege is still relevant in the more recent literature (Alston 1971, Tye 1999, Chalmers 2003, Gerlter 2012). Extrospection, by contrast, is regarded as having to make do with inferences that are thought to be indirect, e.g., because they are based on behavior, verbal report or neuroscientific evidence; it has even been argued that these inferences require a priori principles that do not allow for empirical validation or revision (Goldman 1997, Chalmers 1998, 2000).

Let’s assume for a moment that the underlying assumptions are true: Introspective knowledge is direct and directness constitutes a significant epistemic privilege. Wouldn’t this result in a severe epistemic deficit for extrospective methods? Not really! The reason is that extrospection could then exploit this privileged source of evidence, e.g., by obtaining introspective reports. In fact, this is what extrospective research actually does: It uses introspective reports as an extremely important source of evidence.

But is it true that introspection is direct and privileged? In order to answer this question, we first need to spell out what directness means. We suggest understanding directness in a comparative sense, as it is completely unclear whether any epistemic mode is direct in an absolute sense. Arguably, even the directness of introspection might be questioned, given that introspective knowledge is mediated by a host of cognitive processes and affected by contextual factors, background knowledge, etc. (Sellars 1956, Sperling 1960, Nisbett and Wilson 1977, Schwartzgeb 2002b, Block 2007, Pronin 2009, Michel 2017). Still, it can be argued that extrospection is comparatively less direct than introspection and even less direct than standard scientific methods because, unlike the latter, it relies on some proxy. For example, while a patient may directly know that they are in pain, a physician will have to indirectly infer the patient’s pain from their behavior. In the same vein, we would say that a person who knows about a murder from mere hearsay has less direct
knowledge of the crime than an eye-witness; or that a scholar who has taken the utterance of an author from an encyclopedia has less direct knowledge than someone who has read the author’s original text herself.

What is compared in all these cases are relationships between a fact (e.g. a pain state, a murder, or the utterance of an author) and a person who makes a claim or acquires knowledge of this fact. And the crucial difference between the comparatively direct and the indirect relationships seems to be that the more indirect relationships include intermediaries (pain-behavior, hearsay, encyclopedia) that can substantially affect the transmission from fact to observer, but play no role in the more direct relationships.

Drawing on Alston’s (1971) analysis of “epistemic immediacy,” we can then say that a claim A about a specific fact, e.g., a pain state, a murder, or an utterance is (comparatively) more direct than another claim B, if B depends on intermediaries (pain-behavior, hearsay, encyclopedia) that play no role for A. It would follow then that extrospective claims about a mental state like pain are less direct than introspective claims about the same mental state because the extrospective claims depend on intermediaries like behavioral observation, physiological data, or neuroscientific data, which play no role in introspective claims.

Below we will further distinguish between pre- and post-experiential intermediaries: Pre-experiential intermediaries are factors that are located upstream of the experience, e.g. in the perceptual system and thus affect the experience itself. By contrast, post-experiential intermediaries are located downstream of the experience, so affecting e.g. the judgment about the experience or the behavioral response, without changing the experience itself.

3.3. The objection from indirectness

Let us now turn to the first objection. Saying that extrospection is epistemically deficient because it is indirect implies that directness is an epistemic merit. And this assumption sounds reasonable enough. For example, courts seem to have good reasons to prefer eyewitnesses over hearsay because the latter leads to more errors and a loss of information. And the reason it leads to this is that it involves indirectness, that is, the transfer from the direct observer over various intermediate stages to the person providing a second- or third-hand report. All these potential errors and the loss of information can be avoided if the eyewitnesses themselves show up in court. Similarly, we would say that taking a quote from an encyclopedia might lead to a loss of information, and maybe even to a completely false attribution, which could be avoided if we have direct access to the original text. In fact, the Latin definition of consciousness/conscientia quoted above is a telling example. Scholars have mistaken it for centuries as a quote from Cicero’s Pro Milone, because rather than reading Cicero’s original text they relied on Latin encyclopedias. Unfortunately, the authors of those encyclopedias didn’t bother reading the original text either, such that a definition from the 16th century could be mistaken until recently for a quote from Cicero (Lamarra and Verde, 2014).

But this line of reasoning doesn’t always apply. For example, if an eyewitness speaks an unknown language, our epistemic situation would improve if we call for a translator, even if we would lose directness. Similar examples from science are easy to find. Take electricity. Like brightness or temperature, electricity was initially “measured” by direct experience. For example, Volta put his fingers into small water bowls connected to the poles of a battery in order to directly sense the electrical current. Today, a voltmeter or an ammeter is used as an intermediary between the electrical current and the observer. Even if this decreases directness, it leads to a significant increase of the accuracy of the measurement. In the same vein, a photodiode in brightness measurement decreases directness compared to direct visual observation. But, again, it nevertheless increases the accuracy of the measuring process in comparison to “direct” observation (Goldman 1999, 251).

If this is true, then indirectness neither sets extrospection apart from other scientific methods, nor does it, by itself, compromise its epistemic value, at least not in principle. The principled objection just doesn’t show whether or not extrospection suffers from a fundamental epistemic deficit. If we want to find this out, we need a more detailed epistemic evaluation of extrospective methods that shows whether epistemic indirectness in the case of extrospection works more like hearsay or more like measuring electricity – which is precisely what we will do in the next section.

The understanding of directness that we have suggested above, may help to give a tentative explanation for the differences between cases of indirectness that tend to improve and those that tend to diminish epistemic access. In the latter case, we typically lack control over the intervening factors. This is particularly obvious in the case of hearsay which seems almost completely beyond our control: We just don’t know how the message was distorted on its way from the original observer to the person making the hearsay-based report. By contrast, measuring instruments have always been seen as paradigms of precision because their designers tried to maintain as much control as possible over these instruments and their results. And one reason why additional intermediaries like electronic circuits may be epistemically beneficial is that they increase control.

It would follow that even if extrospection is indirect, the potential disadvantages of this indirectness could be reduced to the extent that we gain control over the intermediaries.

3.4. The objection from a lack of public access

Before we can turn to the more detailed epistemic assessment of extrospection, the second principled objection needs to be discussed. Extrospection seems to fail a substantial requirement that any scientific method must meet: namely, public access. Any claim that scientists make, and any evidence that is said to support a given hypothesis, has to make reference to publicly accessible facts (Searle 1995, 186 sq., Goldman 1997). But this appears impossible for the objects of extrospection: Mental states are private. In particular, their qualitative character is available only to the subject that experiences them.

This seems to be the heart of Nagel’s dilemma that we referred to in the introduction: On the one hand, we can use publicly available data, e.g., from behavioral or neuroscientific studies, which allow us to meet scientific standards. But then we would miss out
on the specific qualitative character of the mental. On the other hand, we can focus on the qualitative character, but then we are restricted to the subjective, first-person perspective, and thus would violate objective, scientific standards.

This “indisputable fact of privacy” which “makes it impossible for the verbal community to maintain precise contingencies” (Skinner 1971, 191 sq.) has also been one of the main reasons for behaviorist attacks on introspection (Ryle 1949, Carnap 1931, 1932). So how can we provide strong justification for extrospective claims about the qualitative character of the mental, and how can we determine whether or not a certain extrospective claim or a measurement result is correct, particularly if it contradicts a related introspective claim?

The above line of thought is built on a hidden premise that can be questioned, namely, that publicly available evidence must be direct. In fact, what we are lacking is direct third-person access to the mental, but we have already seen that directness is not an epistemic merit per se. Indirect public access, however, mediated by all kinds of methods, does exist: E.g., we do have neuroscientific and behavioral methods and verbal reports – as we have indirect access in the case of electricity, atoms, or black holes – without this being seen as a serious problem by physicists or philosophers of science (Overgaard 2015). To quote Richard Boyd (1980): “Scientific knowledge extends to both the observable and the unobservable features of the world.”

Let’s conclude, then, that both principled objections discussed above are unsuccessful. The indirectness of extrospection does not necessarily translate into a fundamental epistemic deficit and, as a consequence, extrospection draws upon publicly available evidence, even if this evidence is indirect. In other words: Although mental experience is private, knowledge of this experience doesn’t need to be.

But again, because what we have dismissed so far are only principled objections, it may still be true that indirect public access to the mental, even if possible in principle, is so limited and unreliable that it cannot provide the amount of reliability and precision that would be needed for an adequate scientific theory of the mental. Extrospection may still suffer from a fundamental epistemic deficit; we just can’t conclude this on the basis of the principled objections discussed so far.

4. Extrospective measurement is still in its early stages

The suspicion that extrospection does suffer from such a deficit seems to be strongly supported when we look at the tremendous differences that set extrospection apart from standard methods in the natural sciences. How else can we explain the stark contrast between the high accuracy of standard physical methods and the extremely low accuracy of extrospection, much less its failure to capture the qualitative wealth of first person experience? Doesn’t this gap itself show that extrospection suffers from a fundamental deficit – even if the basic theoretical considerations discussed above fail to explain it?

In order to answer this question, we will turn to the second step of our argument: namely, constructing a historical and systematic comparison between physical and extrospective methods. What could such a comparison look like? Chasing down particular studies illustrating either the successes or the failures of extrospective methods will hardly help us to arrive at a reliable result. There is, however, one method that is not only essential for almost any empirical investigation, but can also be seen as representative both for the current state of the art in a given field of research and for its future prospects: measuring. Measuring is essential to almost any empirical method. So if skepticism is right and extrospection suffers from a fundamental epistemic deficit, this should severely affect extrospective measuring and its prospects. And while it is virtually impossible to compare scientific and extrospective methods across the board, a comparison seems feasible if we turn to the relatively limited and well-understood field of measurement techniques. Moreover, historical studies regarding the past development of measurement techniques (Chang 2007, Van Fraassen 1985, Van Fraassen, 2008) allow us to derive some meaningful conclusions regarding the prospects of extrospective measuring.

We will focus on those features of extrospective measurement that seem to set it apart most significantly from measurement techniques in the natural sciences: Its dependency on subjective reports, the problems with quantifying subjective experience, and the skepticism that extrospective results are met with. We will show that physical methods suffered from very similar deficits in the early stages of their development. This observation, in addition to indicating that extrospective measurement may still be in its early stages, enables us to see how extrospective methods might be significantly improved in future, which we will discuss more fully in Section 5 below.

4.1. Subjectivity of extrospective methods

Extrospective measurement today almost always requires introspective reports. Take pain as an example. A typical measuring method like the short form of the McGill Pain Questionnaire depends on introspective reports, e.g., subjects’ ratings of their pain intensity on a visual analogue scale between “no pain” and “worst possible pain” (Melzack 1987, Melzack 1975).

This method counts as an “explicit report” (category i) according to the above distinction, and as _epistemically subjective_ because it depends on the pain-sensitivity of patients or experimental subjects. Unfortunately, even if these questionnaires may be extremely useful in clinical contexts (Ngamkham et al. 2012), their accuracy and validity are low compared to standard physical measuring methods. Moreover, while physical methods can draw on the precise definitions of the _International System of Units_ (BIPM 2019), the rating of the McGill questionnaire is based on a largely intuitive understanding of pain intensity (Correll 2006).

Interestingly, the historical situation in the early stages of physical measurement methods was quite similar. For example, in the first two centuries of photometry, subjective approaches that relied heavily on the judgment and intuition of human observers prevailed, although their accuracy was comparatively low – much like the subjective pain-measuring methods just mentioned. One example of such a low-accuracy, subject-dependent measure is Rumford’s photometer. It required an observer to compare the depth of a shadow cast by a given light source (the measurand) with the shadow of a standard candle made from whale oil. But Rumford’s
detailed specifications notwithstanding, the accuracy of this method was limited. Typically, subjective measurements carried a double-digit relative error, which could reach up to 40%. Still, subjective photometry remained the dominant approach throughout the 19th century. Objective methods with much higher accuracy could establish themselves only after photoelectric techniques became available in the 1920s (Chen 2005).

4.2. Quantitative measuring and the establishment of fixed points

Quantification seems to pose another challenge to contemporary extrospective measuring. Take fMRI methods as an example. MRI normally yields signals in idiosyncratic scanner-dependent and sequence-dependent coordinates. Inferences are drawn based on contrastive comparisons between an experimental condition (participant sees object) and control condition (participant does not see object). Thus, typically, the numerical values in the images cannot be compared across experiments or across scanners. However, recently, quantitative MRI has been developed that might alleviate some of these problems. Nonetheless, this quantification is still macroscopic and not directly interpretable in units of neural activity. Thus, if the aim of neuroimaging is to infer properties of neural activity (such as for example a mean firing rate) not even the basic requirements of quantitative scales are met.

By contrast, the pain questionnaires mentioned above use an ordinal scale. One way to make progress here is to establish fixed points (e.g., “no pain vs. worst possible pain”; see 2.1 above), which would give us an interval scale, but this has turned out to be difficult. Again, the same problems came up in the initial stages of physical measuring. For example, in the early days of thermometry, measuring instruments were not calibrated yet, so scientists could only use one and the same thermoscope to measure, e.g., temperature changes in one room. In order to enable comparisons across different measuring instruments and, eventually, quantitative measurement, scientists searched for fixed points. Candidates included the “most severe winter cold”, the “temperature in a specific observatory cellar in Paris”, and the “boiling temperature of water” (Chang 2007, 10). Making a decision between these candidates raised problems similar to those pain researchers are confronted with today: In order to show that a certain candidate (say, the boiling-temperature of water) really is a fixed point, precise quantitative thermometry was needed – but in order to establish quantitative thermometry, fixed points had to be established in the first place. In Section 5, we will show how a bootstrapping process helped scientists to get out of this dilemma. We will also demonstrate that this strategy applies to extrospective methods as well.

4.3. Skepticism in extrospective and physical measurement

One of the most common challenges to extrospective measurement today is a profound skepticism that first-person states can be captured with objective means at all. Long before Nagel’s now classical “What is it Like to be a Bat?” (Nagel 1974), philosophers have expressed deep skepticism regarding objective proxies for phenomenal experience. In Putnam’s (1965) “Super-Spartan” thought experiment, for example, Super-Spartans are said to feel pain despite being completely devoid of any behavioral disposition related to pain. Conversely, the notorious Zombies (Chalmers 1996) show the same behavior as conscious normals do, but have no phenomenal experience whatsoever.

These thought experiments are normally taken to illustrate conceptual possibilities. They thus require a discussion of the underlying conceptual intuitions and arguments, but this would go far beyond the scope of the present paper. Suffice it to say that strong objections against these skeptical arguments have been made from the early days of the discussion, and continue to be raised in the more recent literature (Shoemaker 1981, Searle 1992, Tye 2007, Pauen 2000, 2006, 2017).

The point we would like to make here, however, is that these skeptical objections do not set extrospective methods apart from other scientific methods, and thus do not indicate that extrospection suffers from a fundamental deficit. In fact, skeptical doubts were common in the initial stages of physical methods as well. For example, similar issues were raised in the early days of temperature- and time-measurement. Skeptics like Haüy questioned whether volume expansion was a reliable proxy for heat. Drawing on the caloric theory, Haüy argued that heating and volume expansion were two independent effects of caloric, a fluid which was then thought of as the substance of heat. Thus, a given quantity of caloric entering an object could result in a high amount of volume expansion combined with a low amount of heat, or vice versa. This was taken to mean that volume expansion could not be a reliable proxy for heat (Chang 2007, 67 sqq.). Another such skeptical worry, arising in discussions of time-measurement, questioned how we can make sure that two subsequent periods of a pendulum have the same duration, if the only way to measure duration is based on pendulum clocks (Van Fraassen 1985, 74 sqq.).

All these doubts have a similar structure: They focus on the relation between measuring proxies (volume expansion, pendulum movement, behavior, neural activity) and the measurand itself (heat, time, conscious experience). According to the skeptics, heat may change in the absence of volume expansion, time may pass faster with constant pendulum movement, and phenomenal experience may change significantly in the absence of any behavioral, functional, or neural difference. This would fundamentally undermine the reliability of the measuring process that relies on the proxy.

In some cases, skepticism turned out to be justified because dissociations actually do happen: The volume of fluids, for example, does not always expand proportionally to the rising temperature. Water does not expand between 0 °C and 4 °C. But the decisive factor in the arbitration of these debates was the availability of multiple sources of evidence – e.g., thermometers with different thermometric substances and, eventually, thermocouples (Mach 1896, 36 sqq., Van Fraassen 2008, 126 sqq.) – which allowed for the detection of these dissociations. In most cases (e.g. time and pendulum movement), new measurement techniques could provide independent evidence that the objections were unfounded. As a consequence, valid and accurate measurement techniques could be established, and skeptical objections lost their force: today, the above objections against physical measuring techniques are almost forgotten.
5. Developing strategies for extrospective measurements

Assuming that the conceptual challenges posed by Super-Spartans and Zombies can be met, we could speculate that skepticism about extrospective measuring might diminish in power and prominence as well, provided that sufficiently valid and accurate extrospective methods can be established. But do we have reasons to expect that they actually will be established? In this third step of our argument we will demonstrate that, yes, the strategy that has proven successful in the development of physical measurement techniques can be applied to extrospective measuring as well. This gives additional support to our main hypothesis that extrospection does not suffer from a fundamental epistemic deficit.

Note that we do not ask for, let alone offer, a positive prediction of future achievements. Rather, we will evaluate arguments and review evidence in order to show that a strategy that has been crucial for the development in other areas of science is available to those working to develop extrospective methods as well.

5.1. Bootstrapping, triangulation and conflict resolution

But what is the strategy that led to the stunning successes of physical measuring? Capitalizing on Chang’s (2007, 39 sqq.) and Van Fraassen’s (2008, 121 sqq.) work, we call this strategy an iterative bootstrapping process, which leads to a stepwise improvement of measurement techniques. This process is typically driven from one iteration to the next by measurement conflicts, which are resolved by inferences to the best explanation, based on multiple sources of evidence, which we describe as triangulation.

5.1.1. Bootstrapping

As Chang and Van Fraassen show, scientists used a bootstrapping strategy that eventually allowed them to overcome the almost hopeless first stages in the development of physical measurement techniques. Scientists engaged in bootstrapping insofar as they employed the imperfect methods and the incomplete knowledge available to them in order to improve their measurement techniques step by step, across multiple iterations. In addition, they capitalized on general progress in technology and scientific knowledge.

At the outset, progress in any individual iteration was minuscule, and every change that was made bore the risk of failure. But occasional setbacks notwithstanding, these iterations did not only lead to more accurate results. They also improved the process that led to these results: measuring methods and our understanding of the measurands thus improving the conditions for further progress. This initiated a self-enhancing development which eventually yielded the extremely accurate methods for measuring exactly defined phenomena that we see today in the natural sciences.

5.1.2. Triangulation

But how were scientists able to find solutions and make progress at all, given that their initial knowledge was sketchy, their methods frail, and the technology available to them left so much to be desired? This is where triangulation comes into play. Triangulation, on our understanding of it, implies two steps. First, scientists collect various kinds of evidence, among them experimental results from their own field, methodological knowledge, and knowledge from the natural sciences in general. Second, they look for a theory or a claim that would best explain the available evidence, that is, they make an inference to the best explanation (Harman 1965, Lipton 2008). One advantage of this method, which has been called “the best science provides” (Williamson and Armour-Garb 2017), is its ability not only to deal with imperfect, fallible evidence, but also to account for the individual strengths and weaknesses of any piece of it.

5.1.3. Conflict resolution

Particularly important drivers of the bootstrapping process were the conflicts that would occur when measurements yielded unexpected or even inconsistent results. Resolving these conflicts was certainly challenging, and at times seemed impossible. Their resolution often called for additional data, new explanations, a modification of measurement techniques, or even general scientific or technological advancements. But if found, solutions could have consequences that would extend far beyond the individual problem at hand. They could lead to systematic improvements of the measurement techniques at issue, or advance scientific understanding of the phenomenon to be measured, thus moving the bootstrapping process one iteration forward.

One such conflict occurred in the early days of thermometry. The first thermometers, or “thermoscopes”, sometimes indicated temperature changes even when there were no reasons whatsoever to believe that the temperature had actually changed. In order to resolve this conflict between measurement results and reasoning about the measurand, additional evidence was required – and found: Evidence about changes in atmospheric pressure. Eventually, an inference to the best explanation led to the conclusion that these changes in atmospheric pressure had affected the measuring process. Thermoscopes used the expansion and contraction of air as a proxy for temperature, and as these instruments were not sealed, an increase in atmospheric pressure had the same effect as a decrease in temperature: The air in the thermoscope would contract. So thermoscopes were in fact barometers as well. As a consequence, scientists sealed thermoscopes so that they were no longer affected by atmospheric pressure (Mach 1986), thus resulting in a systematic methodological improvement in thermometry.

5.2. Extrospective methods

Skeptics could argue that it is exactly this strategy that does not work for extrospective methods. Triangulation and bootstrapping require multiple sources of evidence, but this is, they insist, what we lack when it comes to the mental.
In fact, many authors have argued that introspection is the only reliable source of evidence for first-person experience. Feest, for example, holds that “in the case of introspective data, there is only one instrument available: the human mind” (Feest 2012); and Goldman thinks that there is no way to validate introspective claims, even if they might be false at times (Goldman 1997). In a similar vein, Chalmers has claimed that the interpretation of first-person reports requires pre-experimental principles that neither allow nor call for empirical confirmation or improvement (Chalmers 1998, 2000). Irvine (2019), by contrast, has made the opposite claim. In her view, introspective evidence cannot challenge objective data. She thinks that we should – and actually do – prefer objective over introspective evidence in cases of conflict.

In either case, extrospection would seem to lack the multitude of sources of evidence that is required for triangulation and conflict resolution. So rather than moving forward in the series of iterations required by a bootstrapping process, we would end up in a standoff – either because objective data cannot challenge introspection (Feest, Goldman, Chalmers), or because introspection cannot challenge objective data (Irvine).

More specifically, it would follow, first, that it is impossible in principle to resolve conflicts between introspective and objective evidence in any meaningful way. Let’s call this the standoff problem. Rather, according to Feest, Chalmers, and Goldman, introspection would generally prevail in cases of conflict – while, on Irvine’s view, objective data would have the final say. Second, as extrospective knowledge depends on introspective evidence, it would seem that extrospection measuring can never go beyond the limitations of introspection. We will call this the ceiling problem, since it is based on the assumption that introspection imposes a ceiling on extrospection’s achievements. Either way, extrospection would suffer from a fundamental epistemic deficit.

Here we will argue that both conclusions are unfounded. The idea of a standoff between just two sources of evidence, introspective and objective, is misguided. Claims about the mental can be triangulated much like any other scientific claim, and conflicts between subjective and objective pieces of evidence can be resolved because, in most cases, there are multiple sources of independent evidence available. Although introspective reports are extremely important, particularly in the initial stages of scientific development, they comprise just one of these sources. There are, for example, also neuroscientific, behavioral, and physiological data, as well as insights about psychiatric conditions, available to us. And like elsewhere in science, we accept those inferences that best explain all the available data.

As we will show in the present section, this means, first, that there is no standoff problem: Even conflicts between introspective and extrospective evidence can be resolved in a meaningful way, much like conflicts elsewhere in science. What the resolution is does not depend on a priori principles, but, rather, on the entire body of empirical evidence available. And it may well be that the best available explanation leads us to reject the introspective evidence, provided the objective data are strong enough. If the evidence is ambiguous, we can look for additional data before we draw any conclusions. In any case, triangulation and bootstrapping are possible even in the case of extrospective measuring. Or so we will argue. Then, in Section 6 below, we will address the ceiling problem. We argue that the precision and accuracy of introspective reports does not impose a hard limit on the successes of extrospective measuring, even if the calibration of objective techniques depends on introspective reports.

So let’s first discuss the standoff problem, and demonstrate how multiple sources of evidence can justify an inference to the best explanation that would lead to the rejection of an introspective report. One example comes from Anton’s Syndrome, a rare neurological disorder (Anton 1899, Othman, Lee, and Kini 2019, Maddula, Lutton, and Keegan 2009). This case shows how an inference to the best explanation can lead to the detection of an erroneous first-person report, and thus to a resolution of a conflict between introspective and extrospective evidence.

Patients suffering from Anton’s Syndrome are cortically blind due to bilateral occipital lobe damage, but lack insight into their condition, and so claim that they are able to visually perceive their environment. This leads to a conflict between introspective reports, according to which these patients can see, and objective evidence, which suggests they can’t. If the above-cited authors are right, we would be left with a schematic solution in favor of either introspective (Feest, Goldman, Chalmers) or objective (Irvine) evidence. In any case, the above strategy of bootstrapping and triangulation would be unavailable.

However, according to state-of-the-art neurology, this problem does have a solution. Neurologists think that Anton’s Syndrome (henceforth “AS”) patients are actually blind and their first-person reports are mere confabulation. How do they come to this conclusion? Actually, there are three independent sources of evidence that give them reason to reject AS patients’ introspective reports. First, neuroscientific data show that patients’ visual cortex has sustained severe damage, such that it would seem impossible for them to see anything. Second, patients fail very simple perceptual tests, and indeed display highly unusual behavior, which supports the claim that they are blind. For example, AS patients may fail to react even to drastic changes in lighting conditions, run into objects that are clearly visible, try to walk through walls or closed doors, give obviously false descriptions of physicians sitting in front of them, or extend their hands in the wrong direction when asked for a handshake (Maddula, Lutton, and Keegan 2009, Othman, Lee, and Kini 2019). Third, the conclusion that patients confabulate gets further support from neuroscientific evidence: Lesions in the patients’ visual association cortex are thought to explain the lack of insight into their disease (Das and Naqvi 2020, Carjaval et al. 2012). It is thus an inference to the best explanation, based on multiple sources of evidence, that leads neurologists to the conclusion that AS patients are blind, and that their introspective reports are confabulations.

But maybe the buck doesn’t stop here. After all, an opponent could insist that AS patients, even if they are blind, may have visual experiences that are not based in external visual stimuli. A look at the case studies may provide additional evidence that can help to resolve this issue as well. Most importantly, AS patients do not explicitly refer to their visual experiences, nor do they report anything that might give us reason to think they have visual experiences unrelated to external stimuli. Rather, patients refer to their actual environment, obviously using their non-visual knowledge in order to support their claim that they can see. By contrast, it is hard to believe that these reports are based on e.g. visual hallucinations, as hallucinations tend to be unrelated to one’s environment. So you have to assume that some unknown mechanism coordinates these hallucinations with what the patients happen to know about the
environment – but this does not sound like a good explanation. This view receives additional support from evidence that speech and language areas, if disconnected from their usual input, tend to lead to confabulations (Maddula, Lutton, and Keegan 2009): in the case of AS patients, visual areas would likewise be disconnected from their usual input due to the occipital damage.

So the best explanation seems to be that AS patients are blind and have no visual experiences. But here, as everywhere else in science, even the best explanation may still be false. If someone can make a sufficiently strong case for its falsity, this will create a conflict, which would then call for a resolution. A new triangulation case begins, starting with the search for new evidence that might resolve the conflict – say, new evidence about the neural correlates of visual imagery. Once these correlates have been identified, we could look for them in AS patients. If there are such correlates, this would give reason to believe that our hypothesis is wrong and that patients do have visual experience. This could call for methodological improvements in order to avoid similar errors in the future, thus starting another iteration in the bootstrapping process. Conversely, if the correlates don’t show up in AS patients, the explanation presented here would seem better off, and no methodological improvements would be needed. In either case, triangulation seems to provide us with a strategy for deciding disputes about introspective claims in a meaningful way. And that’s the principle we wanted to demonstrate.

This argument gets additional support if we apply the strategy to a similar case, namely to that of Charles Bonnet syndrome (CBS). The case of CBS shows that, depending on the evidence available for triangulation, the inference to the best explanation may lead us to different conclusions. CBS patients are blind, like AS patients, and they too report visual experiences (Eperjesi and Akbarali 2004); but, contrary to what is assumed in the case of AS patients, scientists take the introspective reports of CBS patients to be veridical. One reason for this is that these patients explicitly refer to their visual experiences in quite detailed and sometimes odd-sounding descriptions that are not related to their environments. This makes it unlikely that patients are just drawing on their general knowledge. Another, related reason that their reports are deemed trustworthy is that these patients show insight into their condition: They know that they are blind and that their visual experiences are hallucinatory.

Again, all these assessments are fallible. This is one of the reasons why we do not commit ourselves to the claims made about the patients in any of these cases or experiments. Rather, we take these studies and the other experiments presented in this paper as mere proofs of principle. In the case of AS and CBS patients, the studies show that we can make differentiated assessments of introspective reports because we have multiple sources of evidence that support an inference to the best explanation, and thus allow for triangulation. Moreover, the examples illustrate that this strategy allows for nuanced judgments, rejecting introspective reports in the case of AS patients (contra Feest, Chalmers and Goldman) and accepting them in the case of CBS patients (contra Irvine).

Binocular rivalry experiments provide another example of triangulation and conflict resolution involving introspective reports, which might lead to a new iteration in the bootstrapping process as well. In these experiments, different stimuli are projected to participants’ left and right eyes, respectively: For example, a green grating moving to the left is projected to their left eyes, while a red grating moving to the right is projected to their right eyes. In the subjects’ actual visual experience, however, just one of these stimuli will be dominant at any given time, because the visual system switches between the two stimuli every now and then.

Typically, subjects are asked to make an explicit report (category i) when the switch from one dominant stimulus to the other occurs. This introspective method has various disadvantages, though, including comparatively low accuracy. As Naber et al. (2011) and Frässl et al. (2014) have shown, an objective measure, namely the Optokinetic Nystagmus, provides an important alternative to these reports, particularly because it is more accurate. The Optokinetic Nystagmus is a swift eye movement that follows the direction of a moving stimulus, e.g., of a car moving from left to right in front of an observer. Interestingly, the Optokinetic Nystagmus also indicates whether the right- or the left-moving stimulus in a binocular rivalry experiment is dominant at any given time.

But how do Frässl et al. know that the Optokinetic Nystagmus is more accurate than first-person reports, that is, how can they decide the conflict between introspective report and objective data (category iii) on the basis of the Optokinetic Nystagmus? In order to decide this question, Frässl et al. added a control condition where they presented only one moving stimulus (e.g., the green grating moving to the left) both to the right and to the left eyes and then switched to the other stimulus (the red grating moving to the right) at a specific point in time (Frässl et al. 2014). As a consequence, the investigators knew exactly when the shift between the two stimuli in subjects’ visual experience would occur. They were then able to compare this particularly reliable knowledge with data both from the Optokinetic Nystagmus and from introspective reports in order to resolve the conflict between them.

It turned out that the Optokinetic Nystagmus gave more precise information than the subjects’ reports, which tended to miss quick shifts between the stimuli. Again, all these claims follow from an inference to the best explanation on the basis of multiple sources of evidence, an inference that leads to a decision in favor of the objective evidence. In addition, the case demonstrates how the accuracy of objective measures, e.g., the Optokinetic Nystagmus, may go beyond the limitations of introspective reports. Keep in mind, though, that all these conclusions might be challenged, thus leading to a call for additional evidence, much like in the case of AS patients.

It might be worth noting that the Optokinetic Nystagmus has since come to be regarded as a potential new paradigm for no-report studies that might move the bootstrapping process another step forward. Whether or not it can do so is subject to fierce controversy, as are the prospects of no-report paradigms in general, on which we will remain neutral here (Michel 2017, Michel and Morales 2019, Tsuchiya et al. 2015, Block 2019).

The three examples above suggest that extrospection does not suffer from a specific kind of standoff problem: Conflicts between introspective and objective evidence can be resolved in a meaningful way. Contra Feest and Irvine, various sources of fallible evidence exist in extrospective studies, enabling us to identify problems and to resolve conflicts even between introspective reports and objective evidence. This shows, contra Goldman, that we can validate and even reject introspective reports, at least in principle.

Much like assertions and hypotheses about any other scientific subject, claims about the mental do not follow from one privileged or even infallible source of evidence. Rather, these claims are based on inferences to the best explanation that try to account for all the evidence available. Moreover, the resolution of conflicts may lead to substantial advances in the development of extrospective
methods, including advances in the interpretation of introspective reports: Thus, contra Chalmers, if we have reasons to believe that participants’ reports are likely to be biased or inaccurate under specific conditions, we can and should control for these conditions in our experiments.

All this gives us strong reasons to believe that the strategies of conflict resolution, triangulation, and continuous bootstrapping, which were crucial for the development of physical measurement techniques, can be employed for the improvement of extrospective measurement as well. And this gives substantial support to our claim that extrospection does not suffer from a fundamental epistemic deficit.

6. Extrospective measurement

But if there are no basic roadblocks preventing extrospective measurement from improving in roughly the way that physical measurement did, why is there such a huge gap between current, state-of-the-art extrospective measurement and physical measuring techniques? Why is extrospective measurement still in its early stages? Conversely, if extrospective measuring can be enhanced by the strategies that have proved successful elsewhere in science, are there any positive real-world examples that can corroborate this claim? In this fourth step of our argument, we will address these questions. After providing a brief overview of report-based methods like psychophysics, we describe recent developments in pain measuring in order to show that triangulation and conflict resolution strategies are already bearing fruit. After discussing the ceiling problem, we will demonstrate that the developmental gap between extrospection and physical measuring techniques can be accounted for by the extreme complexity of brain activity, adding another piece of evidence in support of our claim that extrospection does not suffer from a fundamental epistemic deficit. Finally, we will argue that purely objective methods can provide valid measurements of phenomenal experience, at least in principle.

6.1. Report-Based methods

In the “conceptual clarifications” section above, we drew a distinction between three sorts of data relevant for extrospective methods: (i) Explicit reports that directly express a subjective experience, (ii) implicit reports that can be used to extract indirect information about subjective experiences, and (iii) behavioral, physiological or neuroscientific methods that provide objective evidence about subjective experience.

We have suggested that extrospection can capitalize on introspective data and, in fact, extrospective measurement today almost always depends on introspective reports, explicit or implicit. However, the degree of this dependency differs. In some cases, most typically in cases utilizing psychophysics, each individual measurement requires a report, e.g., a description of the psychological response to a physical stimulus. In other cases, introspective reports are required only during the development of an objective method which, once established, does not require further reports. For example, recent approaches in pain measuring, like the Neurologic Pain Signature (NPS), use reports only for the calibration during the training- and test-phases of an objective fMRI-based technique, which then can measure first-person experience with brain data only. Somewhere in between these two extremes are studies focusing on metacognition, or that use the Perceptual Awareness Scale (PAS), approaches that crucially depend on introspective reports but can be connected to objective measures like the Optokinetic Nystagmus or Visual Awareness Negativity.

While the accuracy of report-based methods is obviously constrained by the epistemic limitations of introspective reports, there are still ways to improve them. For example, standardizing responses in questionnaires can help to reduce biases, and systematic variations of stimuli or behavioral measures can be used as control variables. It will turn out that even the order of questions can affect the quality of participants’ responses.

In any case, the interpretation of introspective reports raises problems of its own. Imagine that two participants give different reports upon being presented with the same stimulus. Can we explain the difference by appeal to a difference in the participants’ experiences, or should we assume a difference in their judgments about these experiences? Remember Dennett’s (1988) famous “coffee taster” thought experiment: Two coffee tasters, Chase and Sanborn, report that they no longer like the taste of their company’s coffee. But while Sanborn says that it is his experience that has changed, Chase holds that his judgment, not the experience, is the source of the difference between his past and present preferences.

The underlying problem is quite familiar from the literature on report-based methods. One example comes from discussion of the response bias (Klatzky and Erdelyi 1985). Again, two participants are presented with the same perceptual stimulus, but only one of them reports having perceived it. Two interpretations are possible: Either the difference is perceptual, such that only one participant has perceived the stimulus, or participants may have simply differed in their report criterion, such that the stimulus only met the threshold of the person with the less demanding criterion.

A similar ambiguity affects the interpretation of Solomon Asch’s (1955) famous social conformity experiments. Asch could show that participants’ performance in an extremely simple perceptual comparison task decreased significantly once all other “participants” (who were actually confidants of the experimenter) gave the same – obviously wrong – response. Again the question is how, exactly, the participants’ performance was affected: Was it just the participants’ judgments of the perceived stimuli that were altered, or did the social pressure from the other “participants” rather alter the perception of the stimuli itself?

A final example is provided by placebo analgesia experiments (Shevlin and Friesen 2019, Gligorov 2017). Converging evidence shows that reports about the felt intensity of one and the same pain stimulus can differ, depending on the participants’ beliefs and expectations, e.g., whether they think they have been administered a painkiller or how expensive they think the painkiller was, even if the painkiller was identical in both conditions. Again, the question is whether these reports indicate a change in pain experience or only a change in the judgment about it.
These ambiguities seem to show that the indirectness of extrospective evidence does indeed lead to a fundamental epistemic deficit. But a more careful analysis reveals that this is not the case: Even if the questions we have considered are difficult to answer, it does not follow that they are unsolvable in principle. Rather, we are confronted with a typical conflict of two interpretations which calls for a resolution.

In the case of the response bias, psychologists have suggested mathematical models that are supposed to control for this bias (Green & Swets, 1966) even in metacognition studies (Sherman, Seth, and Barrett 2018).

In the other cases, the account of directness suggested above (1.2) can help to describe the conflict a bit more precisely. At the most basic level, the issue is whether pre- or post-experiential factors account for the difference in response. In the first case, we would assume that a change in, e.g., the perceptual system leads to a change in the experience, while in the second case, the change could come from the judgment about the experience or the subsequent behavioral response, without affecting the experience itself.

Neuroscientific studies may speak to this question. For example, in the case of Solomon Asch’s experiment, neuroscientific evidence has been claimed to indicate that the main difference between participants is pre-experiential, and therefore affects their experiences, not their judgments (Trautmann-Lengsfeld and Herrmann 2014, 2013, Berns et al. 2005).

A similar picture emerges from studies on placebo analgesia: Participant’s expectations and beliefs e.g. about the pain killer activate exactly those brain mechanisms and endogenous opiates that are thought to play an important role in the control of felt pain intensity in general, but seem to be independent from judgment formation (Shevlin and Friesen 2019, Gilgorov 2017). These findings support the more general claim that the content of cognitive states like beliefs or desires may affect pain experience (“cognitive penetration”) even if it is not completely clear that the same holds for perception (Firestone and Scholl 2016).

We concede that the evidence particularly regarding the Solomon Asch experiments is anything but conclusive. However, the studies do show that the question is not as mysterious as Dennett seems to suggest; rather, serious scientific work can help us to make progress regarding even these difficult issues – although there may still be a long way to go.

In order to further illustrate prospects and problems of report-based methods, we will now give a very brief overview of the most relevant ways to collect subjective data in psychophysics, metacognition studies, and experiments using the Perceptual Awareness Scale, before we turn to a more extended discussion of objective measurement techniques, particularly in the case of pain measurement, in the next subsection (6.2).

6.1.1. Psychophysics

While psychophysical measuring directly depends on introspective reports, it does so in a highly systematic way (Fechner, 1966/1860). One typical method of extracting quantitative information about perceived stimulus intensity from verbal reports takes the Just Noticeable Difference (JND) as a unit. The just noticeable difference is the smallest stimulus change that participants can detect above chance. The collected reports about stimulus changes are then submitted, together with data about systematic variations of stimulus strength, to a mathematical analysis. In this second step, additional information about the relation between stimulus strength and sensation intensity is extracted. Because this information goes beyond the explicit content conveyed by the participants, these reports count as implicit (category ii).

The method of just noticeable differences has not remained uncontested, though. Stevens (1957), for example, has argued that just noticeable differences are not constant across the entire range of intensities. That is why he has instead proposed that we use Magnitude Estimation to assess the perceived relationships between stimuli of different intensities. Meanwhile, Norwich and Wong (1997) have suggested to resolve the conflict between Fechner’s and Steven’s approach by making adequate modifications to Fechner’s original theory.

More recent psychophysical studies have moved beyond Fechner’s original approach with its focus on the relation between stimulus strength and sensation intensity, in order to give a more extended picture of first-person experience. One example is David Rosenthal’s (2010) Quality Space theory, which provides a systematic description of color experience, similar to Munsell’s three-dimensional model of color vision (Newhall, Nickerson, and Judd 1943), consisting of hue, saturation, and lightness as dimensions. Employing the method of just noticeable differences, Rosenthal constructs a “quality space in which the distance between any two perceptible properties is a function of how many properties between the two the creature can discriminate” (Rosenthal 2010). He thus provides evidence about the structure of first-person color experience that goes far beyond the individual changes in hue, saturation, and brightness, reported by the participants. That is why these reports count as implicit (category ii) as well.

While this approach raises questions of its own, it – like almost any other approach in psychophysics – shows not only how we can turn introspective reports into extrospective knowledge in a systematic way, but also demonstrates how we can extract information from highly standardized first-person reports that goes beyond what is accessible for any individual subject. Keep in mind, though, that psychophysics does not overcome the direct dependence on introspective reports.

6.1.2. Metacognition

Metacognition studies comprise another domain of research in which introspective reports are utilized in order to collect evidence about the mind (Flavell 1979, Fleming and Dolan 2012). Metacognition covers a wide range of introspective attitudes (Frazier, Schwartz, and Metcalfe 2021), including judgments like confidence ratings or judgments of learning (Rhodes 2016, Son and Metcalfe 2005), but also experience-based attitudes like the feeling of knowing (Thomas, Lee, and Hughes 2016), the “tip-of-the-tongue phenomenon,” and déjà-vu experiences (Schwartz and Cleary 2016). Metacognition plays an important role as a source of introspective information about, e.g., one’s own knowledge, but also for self-regulation, as in learning or memory retrieval (Frazier, Schwartz, and Metcalfe 2021, Blummer and Kenton 2014). As some metacognitive attitudes, like the feeling of knowing or the tip-of-the-tongue phenomenon, refer to subconscious states, the use of these attitudes as a source of information about conscious experience in
extrospective research calls for careful interpretation. In any case, metacognitive judgments only allow for limited conclusions about first-person experience.

In a typical metacognition study, participants are requested (1) to do a first-order cognitive task, e.g., memorize items or discriminate stimuli, and then (2) make a metacognitive judgment e.g. about how confident they feel about their performance in the first-order task. This confidence rating is deemed metacognitively accurate to the extent that it reflects the participants’ actual performance.

Confidence ratings are of specific interest in the present context for two reasons. First, they can be seen as implicit introspective reports (category ii) that provide some information about the experience associated with the first-order task, e.g., the strength or clarity of the perception (Bang and Fleming 2018) – even if any interpretation of these data has to be made with account that perception, like cognition, may be unconscious, and confidence judgments may draw on additional information as well (see below).

Second, metacognition experiments allow for a certain amount of behavioral control over the introspective report: High degrees of metacognitive accuracy or predictiveness (Blummer and Kenton 2014, Rhodes 2016) may give us reason to believe that the metacognitive judgment does provide information about the first-order state and the related experience. This conclusion gets additional support from the fact that adults’ confidence ratings reflect their actual performance much better than the ratings of small children do. This seems to indicate that there is internal information available to subjects, and that training may improve access to this information (Fleming and Dolan 2012, Flavell 1979).

Third, metacognition studies can improve our understanding of the cognitive and neural mechanisms underlying introspective reports. PFC activity, particularly in the rostrolateral PFC (Fleming and Dolan 2012) and activity in the perigenual anterior cingulate cortex (Bang and Fleming 2018) but also activity in the ventral striatum (Hebart et al. 2016) has been associated with confidence ratings. Moreover, there are studies showing that motor information (speed of button press) (Pereira et al. 2020) and information regarding the fluency of first-order responses provide important sources of evidence for confidence ratings. As a consequence, the accuracy of metacognitive judgments can be improved if the judgments are made after completion of the first-order response (Siedlecka, Koculak, and Paulewicz 2020, Siedlecka et al. 2019, Pereira et al. 2020).

Apart from showing that introspective reports are not necessarily direct, these findings demonstrate how a better understanding of the mechanisms underlying introspection can improve our control over intermediate factors, which result from the indirectness of extrospection.

All in all, metacognition studies can provide information about mental states, perceptual evidence, self-related judgments, and the underlying psychological and neural mechanisms, and they also allow for a certain amount of behavioral control. It is however difficult to disentangle the information directly relevant for first-person experience from information about perceptual and cognitive states that need not be conscious.

6.1.3. The perceptual awareness scale (PAS)

An approach that is more directly related to measuring subjective experience, particularly in perception, is the Perceptual Awareness Scale (PAS) (Ramsey and Overgaard 2004, Sandberg and Overgaard 2015). The perceptual awareness scale allows for extracting information about the awareness of stimuli (e.g., in visual identification tasks) from introspective reports in a highly systematic way. Typically, the PAS grades the strength and clarity of a perceptual experience over four categories, e.g., “no experience”, “weak glimpse”, “almost clear experience”, and “absolutely clear experience.” As the critical information about experience is directly asked for, these reports count as explicit (category i); moreover the PAS counts as an ordinal scale according to Stevens’ classification scheme mentioned in the conceptual clarification above because it imposes an order on the categories (e.g., the four listed above), but does not quantify the differences between them. This assessment is, however, not completely uncontroversial (Sandberg and Overgaard 2015).

The PAS seems to be more sensitive regarding residual forms of awareness. Residual forms of awareness might go undetected if dichotomous scales are used, as they distinguish only between the presence and absence of awareness (Sandberg and Overgaard 2015, Overgaard et al. 2008). For example, using a PAS, several studies provide evidence for residual awareness in the blind field of blindsight patients which cannot be obtained with dichotomous scales. According to these studies, it is this residual awareness that explains the patients’ ability to identify objects in their “blind field” (Mazzi et al. 2019, Overgaard et al. 2008).

Moreover, PAS values can also be connected to purely objective (category iii) data as provided by the Visual Awareness Negativity (VAN) in EEG experiments. Interestingly, the objective VAN measures can even reflect the degree of perceptual awareness, as measured with the PAS. This seems to indicate that the PAS could provide an important stepping stone for the establishment of objective (category iii) measuring methods for subjective experience (Sandberg and Overgaard 2015), even if Andersen’s conclusions regarding levels of consciousness have been criticized (Michel 2018).

Finally, the PAS allows for control conditions similar to those of metacognition experiments. For instance, experimenters can look for correlations between actual performance in perceptual identification tasks and PAS scores, even if there are potential methodological problems at issue here as well.

Due to its sensitivity and ability to solicit explicit information about subjective experience in a highly systematic and adaptive way, but also because it classifies the strength of subjective experience in a manner that connects this with objective measures, the PAS can be seen as an important addition to the extrospective toolbox, and might be particularly useful for the calibration and control of objective measures.
6.2. Pain measurement techniques

Recent approaches in pain measurement provide a particularly telling example of objective measures that try to capture subjective experience. There are several reasons for focusing on pain in these debates. In the philosophical literature, pain has been taken as an iconic case of phenomenal experience ever since Herbert Feigl (1956). Pain is one of Nagel’s (1974) and Putnam’s (1965) primary examples of subjective experience, and it has played a pivotal role in the debate about consciousness ever since (Lewis 1980, Chalmers 1996, Hardcastle 1997). Moreover, pain gives us a comparatively high degree of experimental control: Noxious stimuli quite reliably invoke a feeling of pain, and they do so almost immediately.

Meanwhile there are several initiatives underway, such as the Pain and Interception Imaging Network (Labus et al. 2016) or the Open Pain project (www.openpain.org), that make data from individual pain studies publicly available. Data can then be shared and aggregated across different research groups; moreover, existing paradigms can be retested in different labs and with more heterogeneous populations in order to improve the generalizability of the results (Woo, Chang, et al. 2017, 374). Pain measurement thus provides very good conditions for substantial measurement improvements, and the strategies for developing measurement methods explored above can be applied here more easily than it can to be to generic neuroscience methods, where efforts towards standardization and exchange are less common.

Take for example the Pain Qx System, an extrospective and objective measuring technique (painqx.com) that still awaits FDA approval. The system uses EEG signals from the Pain Matrix (Prichep et al. 2011, Legrain et al. 2011). After the removal of artefacts, the data are submitted to an online analysis, which then yields a pain score. Other EEG-based systems use machine learning techniques, which are said to detect pain with a classification accuracy of almost 93% (Vanneste, Song, and De Ridder 2018).

A particularly promising approach has been developed by Wager et al. (2013). Measuring fMRI activity distributions across pain-specific brain regions with machine learning techniques, the authors identified a Neurologic Pain Signature (NPS). The signature can distinguish acute pain induced by heat, mechanical pressure, and electric shocks (Krishnan et al. 2016) on the one hand, from similar experiences, among them non-painful sensations of warmth, anticipation and recall of pain, social pain induced by lovesickness (Wager et al. 2013), negative emotions (Chang et al. 2015), and vicarious pain (Krishnan et al. 2016), on the other. Both the sensitivity and the specificity of the NPS are typically well above 90%.

Even more interestingly, the system can also consistently distinguish between five degrees of pain, and detects the reduced intensity of pain experience after administration of a painkiller (Wager et al. 2013).

6.3. Strategies for improvement

Most importantly, strategies that have proven successful in the development of physical measurement techniques can be applied to pain measurement, and in several cases have already successfully been applied. For example, scientific progress in our understanding of the brain regions underlying pain, as well as technological improvements in machine learning techniques, have been crucial for the development of objective measurement techniques like the NPS or the Pain Qx System.

In fact, pain measurement shows how we can employ the strategies of triangulation and conflict resolution to the end of further improving existing extrospective methods. So, imagine that, after running into a conflict between objective results from the NPS and introspective reports, we have collected additional data from an independent source of evidence. Let’s assume these data indicate that it was the NPS that has failed. Beyond resolving the conflict, these findings could give us hints as to how to improve the NPS in order to increase its specificity and sensitivity. We would thus be able to increase the match between objective results and first-person reports, thereby moving the bootstrapping process another iteration forward.

In fact, Woo et al. have developed a second pain signature (SIIPS1) for exactly this reason. The SIIPS1 measures cerebral activity that is not accounted for by the NPS, but does contribute to subjective pain experience (Woo, Schmidt, et al. 2017).

In a similar vein, introspective reports seem to indicate systematic differences in pain experience between Black and White Americans, while the NPS does not detect such a difference. Much like in the case of AS patients, we have a conflict between introspective and objective data which – as we want to show – can be resolved in a meaningful way by the strategy under consideration.

Two solutions seem possible: There may be a bias in the training sample of the NPS, which would call for a better representation of Black Americans in the sample (Losin et al. 2020). But it might also turn out that a bias in first-person reports accounts for the difference. Something like this might at least partially explain the difference between Black and White Americans in the study just mentioned. Losin et al. (2020) point out that “there is some evidence that higher pain reports in AA (African Americans) may be a learned behaviour in response to a history of inadequate pain treatment.” In this latter case, we would have to adjust our methodology of collecting introspective reports such that these divergences can be avoided or controlled for. Either way, the solution of the conflict could move the bootstrapping process one step forward.

Of course, there is no guarantee that what seems to be the best explanation at a given time may not be wrong, and that what appears as an “improvement” is not actually a step backward. But then we would have to expect that similar conflicts would reappear at some point, particularly if the error is significant.

Even “objective” pain measurement today still requires subjective data for calibration. But as we have seen above, observational data were necessary in the initial stages of brightness measurement as well. In any case, a sufficiently valid and accurate objective method would represent important progress in certain clinical settings, where patients’ abilities for first-person reports are limited or completely absent (Prichep et al. 2011, Makin 2016), and maybe even in court rooms, where objectivity plays a crucial role (Reardon 2015).
6.3.1. How to address the ceiling problem

All this shows how the bootstrapping strategy can be applied to improve pain measurement. But it leaves the ceiling problem unresolved: As the calibration of pain measurement techniques depends on introspective reports, how can this technique surpass the limitations of introspection?

One obvious way to improve a method almost independently of subjective reports is to systematically manipulate a stimulus as an independent variable, and then determine how a given measuring technique reflects the manipulation. This is, e.g., what Frässlé et al. (2014) did in the study mentioned above, when they manipulated a visual stimulus in order to find out whether the Optokinetic Nystagmus or introspective reports better reflect the stimulus change. As we have seen, the experiment indicated that the objective measure was superior to the introspective report.

In a somewhat similar fashion, Ma et al. (2016) used both introspective reports and the NPS to detect differences in the intensity of pain between an experimental group that was administered a pain killer and a placebo group. While the effect of the pain killer showed up both in the introspective reports and in the NPS data, it was significant only in the latter case, giving some reason to believe that the NPS could be more sensitive to this difference than introspective reports. Without question, this interpretation calls for additional evidence, but if corroborated it would provide another example that the bootstrapping strategy can help us overcome the ceiling problem.

But even if additional data would not support our interpretation, the above considerations and the experimental results together show that the triangulation strategy, with its appeal to multiple sources of evidence, can when combined with specific manipulations of stimuli enable scientists to develop objective measuring methods whose accuracy improves upon that of introspective reports.

6.4. Complexity explains the developmental delay

But even with all this in mind, there are still stark differences in accuracy between extrospective and physical measurement techniques, and these differences call for an explanation. We have already argued that the reason for this is not the first-person character of mental experience, but rather the fact that extrospective measuring is still in its early stages of development. But how can we explain this developmental delay? Here we will argue that the explanation derives from the high degree of complexity that distinguishes, e.g., the brain activity underlying experiences like pain from comparatively simple physical measurands like temperature, weight, or brightness. This complexity becomes apparent when we compare the low number of independent variables affecting, say, the temperature of water with the high number of variables affecting a given activity state of the neural networks underlying pain.

Our hypothesis that it is this complexity and not the first-person character of the mental that accounts for the developmental delay can be tested, at least in principle. Our hypothesis predicts that an analogous developmental delay should be found in measurement techniques for similarly complex phenomena that have no first-person character. On the standard view, by contrast, one would expect that the delay should be smaller or even absent if non-first-person phenomena are measured under ceteris paribus conditions. In order to test these predictions, we can compare extrospective measurements of pain with the measurements of non-first-person brain states that use machine-learning techniques as well.

Fortunately, Woo et al. (2017) have published a meta-analysis that does exactly this: It compares the classification accuracy of measurement techniques for pain and for neurological disorders like Alzheimer’s, ADHD, Autism, or Parkinson’s Disease. In these latter cases, first-person experience plays a minor role, if any, since these disorders can be identified by behavioral or physiological symptoms.

Most importantly, Woo’s meta-analysis clearly supports our hypothesis regarding the developmental delay: Contrary to what the standard view would predict, the classification accuracy in the case of first-person phenomena is not significantly lower than that of objective phenomena like Alzheimer’s, ADHD, Autism, or Parkinson. The average classification accuracy across studies in all these cases is roughly between 80% and 90% (Woo, Chang, et al. 2017).

This is not final proof of anything, but it is another piece of evidence that speaks against the idea that extrospection suffers from a fundamental epistemic deficit. Apparently, one of the main reasons for the developmental delay in all these cases is the complexity of brain states, whether or not these states realize first-person phenomena. Note that our hypothesis would also predict that the eventual degree of accuracy that we can expect in extrospective measuring will be on par with that of other similarly complex measurement techniques. So we do not assume that extrospective measuring will ever reach the degree of accuracy that can be seen in measurement methods for comparatively simple physical magnitudes like temperature, brightness, weight, or time.

6.5. Phenomenal experience and objective measurement

In closing, we want to briefly address an important problem that has been raised in the introduction. So far, we have discussed just one horn of the Nagelian Dilemma mentioned at the beginning of this paper. If what we have said is true, then there is no difference in principle between extrospective measuring, on the one hand, and physical measuring, on the other, at least as far as accuracy is concerned – particularly if we account for the stage of development and the complexity of the measurands. But it might still be doubted that the same holds for the validity of extrospective methods as well, that is, that these methods really capture the subjective phenomenon at issue: Pain experience, for example, and not just pain-related brain activity or behavior. If they don’t, this would strongly support the claim that extrospective measuring suffers from a fundamental deficit, because it fails a basic epistemic requirement. Moreover, it might also support Dennett’s (1988) skepticism regarding qualia and phenomenal experience.

The underlying question is a familiar issue in psychological testing. Validity, and more precisely, construct validity, is the central concept that denotes the ability of a psychological test to measure what it is intended to measure, namely the phenomenon or construct
under investigation (Irving 2018, chap. 1). This problem is not restricted to extrospective methods: Intelligence tests, for example, also have to make sure that they really test intelligence and not, say, learning achievements or routine behavior.

Let’s imagine that a measuring device based on the NPS indicates that a subject’s pain intensity reaches 9.54 on a ten-point scale. What are the requirements for this result to be a valid measure of phenomenal experience of pain intensity? In order to answer this question, we will take pain intensity as a psychological construct. We can then apply the concept of construct validity in order to determine whether the NPS scores provide a valid measure for the intensity of pain experience.

Construct validity calls, first, for a precise understanding or even a definition of the phenomenon to be tested (Irving 2018, Messick 1995). Thus, our first question will be whether the definition of pain intensity underlying the NPS captures the related phenomenal experience. Second, it has to be determined, based on this definition, whether the test measures relevant and representative aspects of the phenomenon at issue (Messick 1995). Our second question would thus be whether the NPS’s measuring method really captures the phenomenon of pain intensity as it has been defined in the first step. Third and finally, Messick (1995, 741) stresses that validity is not just a property of the test procedure itself but also a matter of the interpretation of the results. That is why our third question will be whether an NPS score of 9.54 lends itself to an interpretation by the addressee of the result that does justice to the subject’s phenomenal experience.

Regarding the first question, the understanding of pain underlying the NPS follows the official definition by the International Association for the Study of Pain. This definition tries to do justice to phenomenal experience, as it explicitly describes pain as “an unpleasant sensory and emotional experience” (Ma et al. 2016). Accordingly, the McGill pain questionnaire includes almost 80 phenomenal descriptors like “flickering, quivering, pulsing, hot, burning, pinching” (Melzack 1985, McMahon et al. 2013, Chap. 21, Melzack 2005).

It seems suggestive to understand these descriptors as phenomenal concepts (Loar 1990, Chalmers 1996, Papineau 1998, Tye 1999, Buekens 2001, Levine 2001, Chalmers 2005). According to Loar, phenomenal concepts are “recognitional concepts” (Loar 1990) that capture phenomenal experience. Mastering the phenomenal concept “Flickering Pain” will enable you to recognize your current pain experience with a template that is based on your previous experiences of flickering pain (Papineau 1998). Conversely, you can use that very same template to make or understand pain ascriptions to others (Loar 1990). While there is no guarantee that the understanding of pain underlying the NPS or the related descriptors are equivalent to phenomenal concepts in the strict sense, it seems obvious that they do refer to the related phenomenal experience.

Regarding the second question, it seems obvious that the NPS uses only indirect evidence. But as we have already seen, directness is no epistemic virtue in itself, and many physical measurement techniques are indirect as well. Still, it would be absurd to claim that voltimeters measure the Lorentz-force that the device’s magnet exerts on the coil, rather than the voltage of the current. Assuming that measuring is a systematic form of “information-gathering,” (Van Fraassen 2008, 157) this is easy to justify, as long as the proxy provides us with reliable and accurate information about the measurand – in this case, about the intensity of pain as an unpleasant sensory and emotional experience.

The NPS tries to capture exactly this. Most importantly, it draws on first-person reports in order to calibrate the response of the signature. Moreover, divergences between first-person reports and objective measurements are seen as conflicts that call for a resolution – as we have shown above in Losin’s analysis of ethnic differences in pain experience, which explicitly makes use of “unpleasantness ratings” (Losin et al. 2020). It would follow, then, that the NPS does capture the subjective aspect of pain as defined in the first step, individual failures notwithstanding.

As far as the third question is concerned, let’s assume for the sake of argument that the definition of the phenomenon really captures the phenomenal experience of pain, and that the NPS methodology does justice to this definition. In this case, a person who masters the phenomenal concept “Pain” and learns that a patient’s NPS score is 9.54 should be able to understand that the patient does have a terrible pain experience. This becomes clear if we come back once again to phenomenal concepts. As we have seen, these concepts can be used to understand pain ascriptions to others (Loar 1990). This does happen when the NPS indicates that a patient’s pain intensity reaches 9.54 points on a ten-point-scale. In this case, the phenomenal concept “Pain” will enable you to get a fairly clear idea that the subject is having a terribly intense pain experience.

Of course, a pain intensity score of 9.54 gives us only a somewhat standardized piece of information about one small – but important – aspect of pain phenomenology. This, however, holds for physical methods as well. Measuring the strength of an earthquake on the Richter scale won’t give you the full picture of the earthquake’s physical properties either. If you want the full picture of an earthquake or a pain experience, then a scientific measurement may not be the right thing for you to pursue in the first place. On the other hand, the development of measurement techniques may encourage the emergence of additional or more fine-grained categories, as it in fact did, e.g., in the development of brightness measurement. Having started with an intuitive idea of brightness two hundred years ago, scientists now differentiate between (1) the complete luminous energy emitted by a light-source; (2) the brightness of a surface; and (3) the luminous intensity emitted by a light-source in a certain angle (BIPM 2019).

Accordingly, future generations of scientists may try to develop more fine-grained signatures that distinguish different aspects of pain phenomenology. Two obvious candidates are the sensory and affective aspects of pain. Affective pain stands for the aversive aspect of pain experience, which is absent in, e.g., pain asymbolia. Patients suffering from this disorder may even laugh while they experience pain (Ramachandran, 1998). Sensory pain, by contrast, causes one to feel the sort and place of the tissue damage. These two aspects can already be dissociated both on the phenomenological and on the neural levels, and scientists have been able to make subjects feel one aspect of pain without the other (Rainville et al. 1997, Gracely, Dubner, and McGrath 1982). All this gives us reasons to assume that future pain signatures should be able to distinguish between those aspects of pain experience, at least in principle. Note also that such a development could lead to a differentiation in our concept of pain that goes beyond the vernacular notion that we currently use.
7. Outlook

One of the main reasons for philosophers’ interest in extrospection is the problem of consciousness. But it should be clear that, taken by itself, even a perfect system of extrospective measurement would not solve the problems that confront us with respect to understanding and explaining consciousness. Measurement is not explanation. But it can contribute to such an endeavor in at least two ways.

First, it can help to clarify our idea, if not our concept, of the explanandum. In fact, taking Thomas Nagel’s work as an example again, it seems at least questionable whether the “mental life” of a person, or the “what-it’s-likeness” of an experience are sufficiently well defined to allow for a substantive explanation (Dennett 1988). In other words: One of the reasons that explaining these phenomena can seem intractable might be that the above metaphors do not give us a sufficiently clear idea of the explanandum in the first place.

A focus on the measuring paradigm in general, and on pain measuring in particular, might help us make progress on this issue. The main reason for this is that pain research and pain measurement techniques can help to clarify our understanding of the target phenomenon, e.g., because they force and help us to better distinguish between pain and related phenomena, or, on the next level, between different aspects of pain. Rather than talking about and trying to explain the “what-it’s-likeness” of an experience in general, we would then talk about and try to explain pain experience in particular, or, even more precisely, affective or sensory pain – which seems to be a more clearly defined task.

Second, if we are able to measure a specific mental experience with an accurate and valid method, we can make controlled interventions in order to find out whether and to what extent a specific sort of neural activity affects the experience. This can help us to improve our understanding of the relation between the underlying neural mechanisms and the experience in question, thus significantly improving our ability to develop an adequate explanation. This is, by the way, not science fiction: As we have seen, recent progress in pain research shows that the NPS is already used in exactly this way, e.g., for investigating psychological (Woo, Schmidt, et al., 2017) or genetic factors (Ma et al. 2016) contributing to pain experience.

Thus, even if there is still a long way to go from measuring the mental to explaining it, there are strong reasons to believe that measuring methods can make a significant, though indirect, contribution to an explanation of consciousness. Most importantly, the work of developing more successful measurement techniques would put us in a position to apply those methods of investigation that have already proven successful elsewhere in science. This suggests once more that consciousness research is not “special” in any substantial sense of the word – it is just awfully complex.

CRediT authorship contribution statement

Michael Pauen: Conceptualization, Funding acquisition, Investigation, Methodology, Writing - original draft, Writing - review & editing. John-Dylan Haynes: Funding acquisition, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

For helpful comments, discussions and suggestions we would like to thank Reinhard Brandt, Dimitri Coelho Mollo, Elisa Filevich, Reinhard Feibel, Sascha Benjamin Fink, Anja Hartmann, Adrian Kind, Dirk Mende, Dominik Perler, Sera Schwarz, Eric Schwitzgebel, Regine Rørstad Torbjørnsen, Jens Schneider-Mergener, Annette Winkelmann, Ali Yasar, the members of our Research Training Group “Extrospection” at the Berlin School of Mind and Brain, the members of our fellow RTG “Situated Cognition” in Bochum and Osnabrück and two anonymous reviewers.

Funding

John Dylan Haynes and Michael Pauen are funded by the Deutsche Forschungsgemeinschaft project number 337619223/RTG 2386 Extrospection and project number 390523135/EXC 2002/1 Science of Intelligence.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2021.103106.

References

Alston, W. (1971). Varieties of Privileged Access. American Philosophical Quarterly, 8, 223-241.
