



Problems of empirical solutions to the theory-ladenness of observation

Themistoklis Pantazakos¹ 

Received: 20 February 2021 / Accepted: 11 August 2021 / Published online: 18 August 2021
© The Author(s) 2021

Abstract

Recent years have seen enticing empirical approaches to solving the epistemological problem of the theory-ladenness of observation. I group these approaches in two categories according to their method of choice: testing and refereeing. I argue that none deliver what friends of theory-neutrality want them to. Testing does not work because both evidence from cognitive neuroscience and perceptual pluralism independently invalidate the existence of a common observation core. Refereeing does not work because it treats theory-ladenness as a kind of superficial, removable bias. Even if such treatment is plausible, there is likely no method to ascertain that effects of this bias are not present. More importantly, evidence from cognitive neuroscience suggests that a deeper, likely irremovable kind of theory-ladenness lies within the perceptual modules.

Keywords Theory-ladenness · Empirical testing · Perceptual pluralism · Philosophy of cognitive science

1 Introduction

Observation reports are widely considered to be the primary tools for evaluating scientific theories, whether on their own merit or comparatively. For those who harbour realist aspirations about science, observation reports are approximately genuine recordings of the world. The realist idea is that we juxtapose scientific theories with observation reports, decide which theories hold up to scrutiny, and then attempt to improve or replace those that do not. In so doing, our scientific descriptions of the world increasingly identify with the world itself. Yet, the theory-ladenness of observation (Bogen, 2009 for an overview), one of the most distinguished theories in post-positivist philosophy of science, claims that observation reports are influenced by factors owed

✉ Themistoklis Pantazakos
tpantz@gmail.com; ucrhtpa@ucl.ac.uk

¹ Department of Science and Technology Studies, University College London, London, UK

to the observer rather than what is observed. This would amount to a readily evident epistemological problem for the realist: in lack of undistorted observation reports, it is hard to imagine any other candidate suitable to facilitate science's convergence towards veridicality. If already held theories influence what we observe, we end up comparing scientific theories not with the world but, at least to an extent, with other theories.

The theory-ladenness of observation debate surfaced in the late 1950s and has carved a rich philosophical trajectory since. Heidelberger (2003) examines foundational works on the issue by Hanson and Kuhn, and finds that the two scholars underlined different causes for the putative impurity of observation reports. Hanson believed that causal considerations invite the influence of theory on observation, while Kuhn located the effect of theory earlier in the observation circuit, writing that “something like a paradigm is prerequisite to perception itself. What a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see” (Kuhn, 1962/1970, p. 113). In contemporary philosophy of science, the theory-ladenness of observation is an umbrella term for a variety of positions. Votsis (2020, p. 1449) notes that the current thesis has expanded to include way more than theories into factors influencing observation. Some such factors are: linguistic frameworks; conceptual schemes; prior beliefs; the architecture of perceptual systems. Moreover, at the other end of the ladenness mechanism, it is now not just observation reports that are thought to be acted on by the above factors, but several alike elements. Chiefly: sense data; percepts; perceptions; empirical data. To organise this web of connections, Votsis (ibid.) suggests modelling the theory-ladenness of observation thesis as a collection of two-variable, input–output functions. The inputs consist of the factors affecting observation; the outputs form the set of elements that these factors act on. Votsis' model may be expressed as follows:

Theory ladenness of <OUTPUT> by <INPUT>: The contents of <OUTPUT> are a function of¹ <INPUT>.

In a forthcoming work, following research in cognitive neuroscience and its philosophy (e.g. Raftopoulos, 2015), I suggest organising the above functions' outputs as five linked stages of the observation process: sensation; perception; observation; data; phenomena. Sensation regards light intensity computation and the production of the retinal image; perception regards the individuation of the scene into unidentified objects; observation regards ascription of meaning to the surveyed scene. Data are organised, temporally successive observations; phenomena are causal descriptions of data. A strong advantage of this model, I claim, is that it systematizes the theory-ladenness literature, with most of its theses expressible as token cases of the general form function. Examples:

- The contents of sensation are a function of the neural architecture of the human sensory apparatus (as in Quartz & Sejnowski, 1997).
- The contents of perception are a function of priming (as in Brewer, 2015).

¹ Here and below meaning: ‘among other things of’, not ‘solely of’. Save for the very extreme, there is no theory-ladenness thesis contending that the outputs of any observational stage are owed exclusively to a influencing factor.

- The contents of perception are *not* a function of cognition (as in Fodor, 1984; Raftopoulos, 2019).
- The contents of data are a function of likely unconscious racial bias in particular social sciences (as in Gould, 1981).
- The contents of phenomena *may be not* a function of prior beliefs (as in Franklin, 2015).

Apparently, these positions claim quite different things. Despite this, Votsis (2020, p. 1449) insists that compromises in veridicality anywhere along the observation stream is bad news for realist. This is so because any token case of the above model, granted that it is true and that its input is anything but the world itself, serves to undermine the truth-probing ability of scientific theory testing. Further, most of these theses likely feed off each other. Therefore, an infiltration of subjectivity anywhere along the line is unlikely to be undone downstream (*ibid.*). If e.g. a prior belief distorts perception (individuation of the scene into objects), it is hard to see why e.g. observation (ascribing meaning to those objects and their interrelations) would not reflect this very distortion. Votsis, I should note, is a realist himself. Therefore, his realism blueprint is bottom-up: it suggests building from objectivism about sensation and perception to realism about scientific theories.

In this paper, I examine two pro-theory-ladenness theses and proffered ‘solutions’ to them—empirical methods of shutting them down. These theses are:

O: The contents of the observation core² are a function of expertise.

D: The contents of observation are a function of already held theories.

The solution offered for O is testing (Schurz, 2015; Votsis, 2020). It amounts to searching for expertise-influenced elements within the observation core and finding none. The solution offered for D is refereeing (Brewer, 2015; Franklin, 2015). It amounts to bringing an impartial adjudicator within the process of observation. In turn, said referee will be able to cast the effects of already held theories away. I argue that none of these solutions work to banish theory-ladenness and vindicate something like the approximate objectivity of the related observation stages. I consider testing and refereeing individually. I provide counter-arguments specific to each one, as well as some that overlap, seeing as I find the two solutions guilty of committing two common central mistakes, namely to ignore how early in the observation circuit theory-ladenness occurs, and the pluralism involved in the human observing apparatus. Overarching remarks on these issues are provided in the conclusion of the paper.

2 Presentation of testing

As mentioned above, testing comes as an empirical suggestion to invalidate O:

O: The contents of the observation core are a function of expertise.

The latest version of the testing solution is put forward by Votsis (2020) via extending a recommendation by Schurz (2015). The term ‘observation core’ is used to refer

² The ‘observation core’ is one leg of the output of observation. I explicate the concept below.

to an early output of the observation circuit, purportedly untouched by expertise. The distinction is important; had O been referring plainly to ‘observation’, it would have been straightforwardly true. Expertise obviously plays a role in meaning ascription to the scene. We hardly need to conduct an experiment to know that an expert physicist would observe white tracks in a cloud chamber as indications of passing electrons, while a layperson would observe the same scene as something along the lines of “white spots forming lines”. With his testing recommendation, Votsis seeks to uphold that, should the experts be forced to rely on their observational capacities *minus* their theoretical training, they would observe the same things as the laypersons. Therefore, expertise does not modify the contents of the observation core, and O is false. Should this be the case, the anti-relativist could claim that, while observation is theory-laden, the ladenness is introduced only via bridge principles that connect the contents of the observation core to theories proper (Votsis, 2015, pp. 580–581). That is, the falsity of O could serve as a springboard for countering the strong relativist bite of theory-ladenness. The realist project would then not be doomed in the offing, at least not regarding the observation core and its (non-)defilement by expertise.³

The experimental design that accompanies testing goes as follows. I omit the thorough technical details and include the parts that are important for the present argumentation. First, gather a number of experts from a scientific field and an equal number of laypersons with no expertise. Second, ask the experts to select a number of instrument-produced images from their field. Third, ask both the experts and the laypersons to draw faithful, no detail spared reproductions of these pictures. Fourth, digitise the drawn pictures. Fifth, ask everyone to match the digitised pictures with the images originally selected by the scientific experts. Sixth, record and statistically analyse their choices (Votsis, 2020, pp. 1455–1458). In evaluating the results, one should not be interested in whether the test subjects are making the right choices, whichever those may be, but in whether the laypersons’ and the experts’ judgments *converge* towards one another. The question put to the subjects is a “what is like what” type rather than “what do you see here” type, in the hope that such a ‘mix and match’ approach will force the experts to lean on their theoretical background as little as possible, as discourse is more likely to carry with it the import of expertise. Should the experts appear to be making similarity judgments different to the laypersons’, it is reasonable to assume that their theoretical expertise accounts for the emerging discrepancy. On the other hand, and if certain preconditions to be discussed below obtain, convergence would imply that expert training leaves no mark on core observations and that, on some basic level, all subjects observe the same:

To the extent that the classification judgments of experts and non-experts are highly convergent ... it is reasonable to conclude that the two groups recognise the same patterns of features in the images, the drawing and hence the world or at least that any theoretical prejudices related to expertise are kept at bay. (Votsis, 2020, p. 1456)

³ The bibliographically informed reader will no doubt find similarities between Votsis’ approach and Fodor’s (1984) widely influential one. They are quite different, however, in that the former regards the early parts of observation while the latter regards perception. As the model of Sect. 1 shows, the two do not coincide.

As it stands, Votsis' experimental prescription may give rise to a number of protestations. For example, what if the task of matching is too easy and, compared to the observations made in scientific context, uncharacteristically easy at that? That is, what if experts cannot but rely on their expertise only in cases of scenes difficult to parse out (as scientific scenes often are) and the similarity task is just too easy to happen upon this phenomenon? Votsis (2020, p. 1457) wards against this possibility by specifying three different collections of images that take a gradual upturn in the difficulty of distinguishing between each collection's members (judged by expert eyes). This is just one potential objection, of which Votsis considers several, and does so carefully. Among the most interesting are those highlighting that drawing itself may be a theory-laden activity, and that both specifying what kind of similarity between the pictures we are asking for *and* leaving it unspecified may create epistemological problems regarding the jump from convergent experimental results to convergent world observations (Votsis, 2020, pp. 1458–1461). The issue of whether Votsis' answers to these objections actually achieve the intended goal outspans the scope of this paper. Therefore, I will refrain from discussing these objections further, noting here only that Votsis is aware of them and has answers in stock so as to avoid doing an injustice to him.

Votsis' point turns on the hope for convergence of judgment. However, he knows well that convergence *alone* will not do the required realist heavy lifting. The problem may be simply put: how do we know that we do not all agree on the same wrong thing? To cast this possibility away, Votsis (2020, pp. 1461–1464) considers explanations of (potential) convergence other than veridicality, and finds them faulty. Social constructivist explanations are wrong, Votsis contends, for his experimental design leaves no room for social negotiation, as individuals are tested in isolation, having no opportunity to confer about their judgments.⁴ Moreover, even if subjects communicated, communication itself presupposes the veridicality of observation. If one subject does not e.g. produce structured sounds of speech correctly and another subject does not decode them correctly, communication between them is an impossible act. Votsis (2020, p. 1463) finds no reason not to expound from this to the veridicality of experiment-related observation itself.⁵ Following this brief rejection of the social variety of constructivism, Votsis considers its neural counterpart. In this case, the neural constructivist position is that observations converge due to their being products of similarly structured brains. However, the line goes, the outputs of our brains are constructs and thus non-veridical. Votsis (2020, p. 1462–1464) has two ready answers to this. First, as long as similar brains produce same constructs from same stimuli, it follows that distinct elements in the set of perceptual constructs match up to distinct elements in the set of stimuli, and thus such constructs reveal something about the structure of such stimuli. Votsis' argument here is essentially this: suppose that two (or more) subjects with different perceiving apparati construct different percepts from the same stimulus, and give this stimulus different names, say A and B. In so far as the stimulus-percept matching is carried out consistently via the two perceiving apparati,

⁴ It is, I think, improbable that friends of social constructivism would grant that the latter is present only when active social communication obtains. However, I do not explore this objection herein.

⁵ Again, I remain hesitant as to whether this argument works, but the scope of this paper does not allow for further related investigation.

the two subjects will likely conclude that they are using different names for the same thing (Votsis, 2015, pp. 566–571). In other words, if the two subjects agree to ascribe A and B a common term C, they will both point at the same thing when asked where C is in the presence of a C. The argument is of course a derivative of Locke's inverted spectrum argument, intuitively graspable by considering that a person who sees green for red and red for green since birth will have no problem navigating New York's traffic lights. Votsis' second answer to neural constructivism is the usual argument from success: if our constructs were not veridical representations of the world, then it would be some sort of cosmic coincidence that we managed to stick around as long as we did as a species (Votsis, 2020, pp. 1462–1464). The argument from success is always the veridicalist's flagship argument, Votsis contends:

[T]he most powerful argument for the veridicality of observational judgments is, as it always was, the success those judgments confer on our ability to predict, and interact with the world. (Votsis, 2020, p. 1464)

In summary: for Votsis, social and neural constructivist explanations of potential convergence regarding the contents of the observation core are wrong. Such potential convergence is best explained by the veridicality of the observation core. Therefore, if testing brings the necessary convergence to the fore, two inferences that the realist would undoubtedly welcome are corroborated. First, O is false. Second, the contents of the observation core are veridical.

3 Why testing does not work

The recommendation by Votsis building on Schurz is likely the most compelling approach to testing theory-ladenness articulated in recent years. Nonetheless, I do not believe that friends of theory-neutrality should, in the end, look to it for strengthening their rhetorical arsenal. This is for three reasons. First, regardless of what the results of the hypothetical experiment show, research in cognitive neuroscience has already demonstrated that Votsis' observation core is penetrated by expertise in general, and by theories proper in particular after only 150 ms following stimulus onset. Therefore, O is true. Second, even if O is somehow false and the observation core is theory-free from the angle of expertise, it is still laden in another way, important for the veridicality of its contents and thus for realism. The argument here comes from considering cases of perceptual pluralism, which prompt the conclusion that there is not one observation core, but many, populated by different percepts. Moreover, I argue that, since there is no single observation core that is uniquely successful in its employment to navigate the world, we likely have no reason to single out the contents of just one of these cores as veridical. Therefore, objectivism about the observation core is undermined. Third, even if the first two reasons are ignored, it is unclear how one may build from theory-neutrality of the observation core to objectivism about observation (and, much more so, to scientific realism).

3.1 Concepts move in too fast

Votsis does not explicate the ‘observation core’ concept as he uses it in great detail. From what has been said, however, it is clear that he employs the term to refer to something like “what we see before what we know kicks in”—in fact, the whole experimental process is an attempt to establish that such a thing exists and can pave the way to realism. The last 30 years have seen copious attempts at defining such a stage of the observation process in cognitive neuroscience and its philosophy. Fodor (1984) is the contemporary inaugurator of this work, which is today carried forward by Raftopoulos (2001, 2009, 2019) and others. Raftopoulos, as someone fighting on the side of objectivism about perception for the last 20 years, has admitted from the outset (Raftopoulos, 2001, pp. 426–427; 438) that perception is cognitively penetrated after the first 150 ms following stimulus onset, a view he still holds today (Raftopoulos, 2015, pp. 87–92). According to him, information held in the higher cognitive centres like the prefrontal cortex, purportedly the locus of theories proper when located in our brain, starts modulating the content of perception *this* early, mainly by recognizing objects proper. This is just for theories proper; the overall mark of expertise, however, stretches beyond the activation of a set of adopted beliefs. Raftopoulos (2001, p. 443) underlines that perceptual modules are open to long-term rewiring as a result of learning. This is learning integrated in perception, which acts even before semantic influences from theories proper kick in to modify the contents of perception (> 150 ms). Therefore, expertise’s effect is there from the onset regarding perceptual learning, and post-150 ms after the onset regarding the import of theories proper as a result of expertise. Raftopoulos’ conclusions are particularly important here for two reasons. First, his research is widely cited and well-known to be up to date with latest findings in cognitive neuroscience. Second, he is the most influential defender of *objectivism* towards perception within the theory-ladenness debate today, and *even he* has not ventured beyond the first 150 ms in trying to establish such objectivism. Without even having visited the non-objectivist camp (e.g. Churchland, 1988), that is, the available bibliography allows us to aspire to expertise-purged perception only up to the first 150 ms after stimulus onset—roughly the maximum recommended delay for telephone services! Now is a good time for a reminder of O, which Votsis seeks to invalidate:

O: The contents of the observation core are a function of expertise.

That any of Votsis’ experimental tasks are able to be completed under 150 ms is, I take it, self-evidently excluded. Therefore, the input of expertise will almost certainly show up in his experimental results, and O will be shown to be true. *Even if* we exclude perceptual learning and take expertise to mean only the adoption of theories proper (and there is no palpable reason that we should), the above shows that perception is at best purged from expertise’s effects only during perception’s initial 150 ms. This is far less than Votsis means with the term observation core, since his tasks for extracting its contents, like drawing and matching, take far more than 150 ms. One may at this point think to rectify the problem by suggesting that all activities which introduce recognizing objects proper (chiefly drawing) be replaced with other activities that do not. It is, however, far from evident what kind of activities these should be. Since Votsis

purports to construct an experiment that takes similarity judgments as its primary data, subjects *need* to have (minimally) a way in which to report such judgments available to them, and I can think of no such process that takes under 150 ms for completion. That is, even if one is charitable with Votsis and allows e.g. drawing to be something like a placeholder activity, it is unclear what the reporting tool adequately filling this place would be.

Before I move on to my second counter-argument to testing, I would like to explore two possible objections to my first. The first objection goes as follows. It is true that 150 ms elapsing for the experimental tasks to be completed ensures that the application of concepts will have infiltrated the process. However, even within conceptual vocabularies, there is a distinction to be made between plain perceptual concepts and concepts acquired and/or shaped by expertise.⁶ It has been argued (e.g. Raftopoulos, 2015, p. 94; see also my second counter-argument to Votsis below) that the plain application of concepts (i.e. not expertise-imbued ones) is not vicious for veridicality (and for realism), since we all live in the same world, acquire the same basic perceptual concepts, and the veridicality of these concepts is safeguarded via evolution. To this hypothetical objection, I have two retorts. First, it not clear that, because we all live in the same world, we all acquire the same basic perceptual concepts. In fact, extant examples from clinical and non-clinical bibliography demonstrate the existence of multiple sets of perceptual concepts. Equally importantly, none of these sets are uniquely successful, thus rendering the relevance of other sets for veridicality mute. Of course, this point is not self-evidently valid; I argue for it below, in my second counter-argument to Votsis. Therefore, my first counter-argument depends, in this limited sense, on how convincing my second counter-argument to the experimental design in question is. Second, even if we all share the basic framework of perceptual concepts because we all live in the same world, it is unclear how one filters the import of higher cognitive centers, allowing for plain concepts to mould perception, but not expertise-borne-and-shaped ones. As long as the experimental tasks take more than 150 ms to be completed, expertise-related concepts *are* available to the acting circuits. Perception, my point is, is in part not a *guided* process; there are no buttons which one may press to let just one kind of concepts do the recognizing work. That is to say, even if only the import of expertise-related concepts is vicious, there is no reason to suppose that these concepts will be ‘left behind’, since the tasks will take more than 150 ms for completion. In fact, Votsis’ experiment is a protracted attempt to ‘trick’ experts to leave concepts exclusive to them behind. Alas, this proves a task much more difficult than initially imagined.

The second objection that could be levelled against my first counter-argument to Votsis has to do with perceptual learning, the effects of which I took to be damaging for the integrity of the proposed design. A popular view is that the effects of perceptual learning are irrelevant for theory-ladenness because perceptual learning is data-driven, not theory-driven (see Fodor’s and Raftopoulos’ replies to Churchland in Fodor, 1988 and Raftopoulos, 2001, pp. 430–432 respectively). Siegel (2011, pp. 5–6) is clear about this (*italics mine*): “if visual experience is cognitively penetrable, then it is

⁶ The distinction itself appears to me fuzzy, but this is a complicated matter (see e.g. Chang, 2012), and at any rate not one that can be settled here. I thus will be charitable with the supposed objector, and grant that the distinction rightfully obtains.

nomologically possible for two subjects ... to have visual experiences with different contents ... as a result of differences in other *cognitive* ... states". Friends of theory-neutrality focus on the impact of cognitive elements because they hold that proper control of non-cognitive elements makes subjects have the same (veridical) percepts when encountering the same stimuli. Ergo: if it is not a result of cognitive difference, it is immaterial to theory-ladenness. I put forward two problems for this line of thought. First, highlighting that proper tests of theory-ladenness control for stimuli, attention, and perceptual learning, is not able to rescue Votsis' design from my criticism. Votsis' proposed experiment includes laymen and experts, subjects who, by definition, have had at least very different perceptual trainings when considering stimuli relative to the experts' ambit. Therefore, following the above authors, his test is not appropriate for testing theory-ladenness since it intentionally does *not* control for these factors. Second, and more substantially, I disagree with the above authors: that the effects of perceptual learning are not explainable in terms of cognitive states does not make said effects irrelevant to theory-ladenness, even if such learning is data-driven. To realise this, one has to ask: *which* perceptual training is appropriate for shaping an observation core with veridical contents? If one makes the plausible assumption that different Kuhnian paradigms come with different perceptual and attentional trainings, it is reasonable to ask: from which one does the path to veridicality start? The answer to this question seems unstraightforward, or minimally turns on the discovery of a criterion for the relevant appropriateness of paradigms. Notice that both my points are entirely unaffected by the assumption that controlling for non-cognitive elements gives the same percepts,⁷ and by the assumption that just some sets (ideally for the objectivist, one) of these non-cognitive elements are veridicality-conducive. Another way to put the same point is this: we may be able to control for all things non-cognitive, but just which non-cognitive things we should go for is underdetermined. Therefore, perceptual rewiring as a result of training in different paradigms remains relevant to theory-ladenness, not in the sense that it necessarily regards cognitive states, but in the sense that it threatens to undercut the veridicality of the observation core.

3.2 Perceptual pluralism means no observation core commonality

My second counter-argument to Votsis' line departs from perceptual pluralism and does not depend on the first one. As is widely known within perception-meddling professions, human perception consists in fact a multitude with significant variations. Since space is limited, here I restrict to citing a number of perceptual atypicalities, starting with some found in a comprehensive review article by Ffytche et al. (2010). These are: atypical size and object perceptions (metamorphopsia); indiscriminable visual perceptions (gnosianopsia); discriminable visual perceptions without having said perceptions (agnosopsia); perceiving multiple copies of the same subject (polyopia); objects stuck at particular spatial co-ordinates within the visual field (visual perseveration); objects returning to the visual field (delayed palinopsia); a pattern from an object spreading to its surroundings (illusory visual spread). By bringing these examples of perceptual variation to the fore, I am no longer pushing for O, but

⁷ I will disagree with this contention in the next section, but I grant it for now.

for a larger point. To see this, grant that O is false and that, indeed, the observation core is somehow insulated from the input of expertise. Votsis' hope is that there is commonality regarding the contents of the observation core; that minus the effects of expertise, perceptual contents are the same. Then, supposing that commonality is established, Votsis argues for the veridicality of the observation core's contents. Perceptual pluralism means that the contents of the observation cores are not the same across subjects, so the argument building from commonality to veridicality falls through. Take the case of indiscriminable visual perceptions (gnosianopsia) as an example: it is evident that a subject who cannot even discriminate between visual perceptions can neither perform a picture matching task like a subject who can, nor acquire a concept based on visual presentations like a subject with typical perception. There is no Votsian convergence to begin with; no common individuation, parsing, extracting something non-kaleidoscopic from the passing scene is to be found across subjects.

No doubt, the first retort to my second counter-argument on behalf of the objectivist will be that the perceptions I have brought to back it up are somehow problematic as mostly encountered within clinical literature. Indeed, Votsis (2020, p. 1456) writes explicitly that one of the preconditions for partaking of his experiment is "normal visual perception, i.e. no visual impairment". Votsis does not provide an argument for establishing this precondition, leaving us the necessity to speculate as to why. It is reasonable to assume, I think, that Votsis considers these perceptions less successful in coping with the world than the typical, and therefore less truth-warranting. There is however a crucial problem with this position, evident in Votsis' confusing 'normal visual perception' with 'no visual impairment'. Recent scientific literature supports that many perceptual atypicalities do not even imply an accompanying disease or pathology (Ffytche et al., 2010, p. 1280). Blom's (2010) influential treatise of hallucinations, *A Dictionary of Hallucinations* includes numerous atypicalities in perception that do not count as pathological in clinical psychiatry and neurology. To name just a few: negative afterimages; Eigengrau; Haidinger's brushes; flick phosphenes; pressure phosphenes; the motion aftereffect; Moore's lightning streaks; erythropsia; fata morganas; hypnagogia; synaesthesia. Moreover, there exists a strong tradition of otherwise academic (Stanford Neurodiversity Project, n.d.; Pantazakos, 2019a) and political (Kras, 2010; Ripamonti, 2016) argumentation for why neurobiological atypicalities, also regarding irregular perception, should be conceptualized as legitimate differences. Phenomenological reports and literary treatises by people in perceptual atypicalities (e.g. Grandin, 2012; Higashida, 2013) and/or professionals working in the field (e.g. Sacks, 1985/2011, 1995) document how people in atypical perceptions navigate the world just fine. To put it minimally, then, discarding this argument would take establishing the non-veridicality of at least the perceptual atypicalities cited above. If that much is not possible, the contents of the observation core are importantly different, and building from commonality to veridicality has taken a foundational blow.

Suppose, however, that the objectivist wants to push that line, and somehow manages to convincingly cast most cases of perceptual variation as inferior. Evolution, presumably their argument would go, has chosen typical perception for a reason, i.e. *because* it is veridical. Gnosianopsias and other bumps along the road are just expected variations, and it is no coincidence that they are both clinically relevant and somewhat rare; this is exactly because they are non-veridical, creating problems in coping with

the world (though, as just demonstrated, these statements are far from obviously true). What the objectivist would have to deal with then would be cases of atypical perception that are, by extension of this logic, *superior* to the typical.⁸ Such examples abound. Autism spectrum conditions (hereafter ASCs) often come with what is characterised as enhanced perceptual processing (Happé & Frith, 2006; Mottron et al., 2006). Mottron et al., (2009, p. 1385) write that perception as a whole should be viewed as an integral part of the mechanism of savant abilities. The widely acclaimed Enhanced Perceptual Functioning (hereafter EPF) model of explaining the main differences between autistic and non-autistic perceptual processing (Mottron et al., 2006 for a recent version) is rich with examples of what is called superior perception related to ASCs. According to EPF, operations that are executed superiorly by people with ASCs diagnoses can be explained as part of a superior perceptual functioning (Mottron et al., 2006, p. 28). A “primary superiority in perceptual analysis” write Mottron et al., (2006, p. 28), “could possibly underlie ... exceptionally accurate reproduction of surface properties of the world, like 3-D perspective or absolute pitch values in savants”. Overall, the preferred processing of local (versus global) information on behalf of people with ASCs diagnoses, responsible e.g. for their not falling prey to certain illusions (Ropar & Mitchell, 1999), is attributed to a superiority of low-level perceptual processes according to EPF (Mottron et al., 2006, p. 29).

Casting such observation cores as inferior to the typical is obviously an unavailable option since, by virtue of these cores, subjects who possess them perform superiorly in standard perceptual tasks (thus ‘enhanced’ perceptual processing). Do we then have in our hands examples of observation cores that the objectivist cannot but admit as legitimate alternatives to the typical, and which thus threaten to undercut the Votsian argument building from commonality to veridicality even more forcefully than before? This would be too fast. A theme running through the investigation of EPF subjects’ perception is that these subjects construct *more fine-grained* percepts than the standard, something which is clearly not the case with the previous bunch of perceptual variations. Thus, the objectivist could claim, the EPF subjects do not possess observation cores that are genuinely different to the typical. A useful parallel could be drawn here with the case of Eskimos who, due to their natural habitat, have been trained to perceive fine differentiations of snow that e.g. a person who has been born and raised in sub-Saharan African cannot, at least not without the proper training. However, to claim that they are seeing something different when looking at snow, and that this is somehow relevant for theory-ladenness and/or realism, may be taking it too far.⁹ To the extent that our grain-responsible perceptual faculties diverge, we are all picking up different grain-level percepts, but this alone does not legitimize the claim that we are perceiving different things. Nevertheless, notice that to state that different levels of grain do *not* make for different percepts is not self-evidently true either. Extreme differences in granularity can be and have been said (see e.g. Block, 2013) to make for different percepts. Evidently, the matter cannot be settled here. Setting it aside and

⁸ The notion of a non-contextual notion of success and inferiority/superiority regarding perception is in my own view highly problematic. However I do not have the space to pursue this matter here. The point here is that typical perception, if superior to clinical examples of atypical perception, is then inferior to other examples of atypical perception.

⁹ I thank an anonymous reviewer for pointing this possible line of counter-argumentation out.

moving temporarily back to the more restricted frame of Votsis' proposed experiment, it is definitely the case that EPF-related differences are directly relevant to the tasks included therein. The proffered by the EPF primary superiority in perceptual analysis, exceptionally accurate reproduction of surface properties, and preferred local processing of information, are exactly the kinds of things that feature in drawing a faithful copy of an image, and detecting similarities between digitized drawings and originals. It is then plausible to state that Votsis' test will show differences in observation cores between neurologically typical and EPF subjects (though these differences will not be due to expertise).

To take stock, my second counter-argument so far is that both 'inferior' and 'superior' perceptions make a case for the non-commonality of the observation core's contents. 'Inferior' perceptions do evidently produce genuinely different percepts, and at least some are plausibly not inferior at all. 'Superior' perceptions regarding EPF subjects produce more fine-grained percepts. The extent to which this undermines the argument from commonality to veridicality depends on the degree to which these more fine-grained percepts are genuinely different percepts; the extent to which EPF subjects see different things when looking at the same stimuli. Regardless of this larger matter, it is very likely that the perceptual differences in EPF subjects *will* feature in the results of Votsis experiments if the experimental subject pool is appropriately varied, pinpointing differences in the observation core. Further, it is quite surprising to read that Votsis himself, despite posing the "no visual impairment" precondition, does not ascribe much epistemological importance to the inferior-superior distinction in perception (footnote mine):

[O]ne can easily imagine individuals with 'enhanced', as opposed to 'impaired', colour- or face-detection abilities. Such individuals would also deviate from the aforesaid norm. As before, since the veridicality of perception is a point of contention in the philosophical literature, we can put aside any judgments that such individuals are either impaired or enhanced and merely note that differences in sensory physiology have an impact on perceptions. ... [A]ll of these input–output relations¹⁰ threaten the neutrality and truth-probing ability of scientific theory testing. Moreover, note that the threat is even more unified than perhaps first imagined in that the aforesaid relations feed off each other. For if we assume that the outputs, to the extent that they are real and distinct kinds, are linked stages on the path from stimulus to observational reports, then it is not unreasonable to maintain that any change effected early on in that path is not likely to be undone downstream. (Votsis, 2020, p. 1450)

Votsis does not, of course, state that the impact of sensory physiology on the percepts is *enough* to establish the theory-ladenness of the observation core. He grants, however, that extant cases of superior perception are not even necessary to establish eligible examples of perceptual variation. Even *hypothetical* ones, extrapolated from cases of inferior perception, will do. In my previous argumentation, I provided examples of 'inferior' perceptions with different percepts. If Votsis' view obtains, we can

¹⁰ Of which one is the input–output relation between perceptual abilities and percepts, including the case of 'inferior' abilities.

extrapolate from these cases to cases of superior perceptions with genuinely different percepts. Thus, any argument about the inferiority of different percepts would crumble, and the existence of ‘inferior’ perceptions with different percepts would be enough to disrupt the commonality-veridicality argument in its inception. Of course, that Votsis himself is an objectivist granting the eligibility of suchlike extrapolation does not mean that this extrapolation is not likely to face reasonable opposition from other objectivists. Therefore, I am not considering this issue solved by virtue of the present syllogism; I am here only showing what follows should we entertain Votsis’ considerations.

A last note: *even if*, somehow, typical perception is superior to all others and converging towards veridicality, the objectivist’s case about perception is not secured. Basic facts about the phylogenetic and evolutionary history of perception reveal that the human typical perception of our times is but a point on a trajectory (see e.g. Martin & Gordon, 2010). Even if typical perception was somehow singularly successful and evolutionarily converging towards objectivity (and these are quite big ifs), then our evolutionary progeny’s perception should be even more objective than ours. Here is then the question: after which point in this convergence trajectory are we legitimized to call perception veridical? The question seems to have no non-arbitrary answers, thus hampering the singling out of an evolutionary station (conveniently, ours), calling it veridical, and ascribing it powers of objective world-taking.

It is by now evident that my second counter-argument is throughout a neural constructivist one. However, as mentioned early in the previous section, Votsis has addressed, and purports to have done away with, neural constructivism in his original paper. Reminder that his arguments against neural constructivism are (a) that if same constructs follow same stimuli regarding similar brains, then something is learned about said stimuli, and more importantly that (b) we navigate the world impressively well with our constructs, therefore they must be approximately true representations of the world (Votsis, 2020, p. 1464). Point b is, and always has been, by Votsis’ own words, rightfully the flagship argument for veridicality (ibid.). What I have argued above, I believe, shows that one cannot shrug off neural constructivism simply by appealing to a and b. If my relevant arguments obtain, different brains do not just produce inverted percepts; they produce different individuations and parsings of the passing scene, and appealing to the inverted spectrum argument is inadequate to address the issue. Moreover, the human species creates a manifold of sets of perceptual constructs, of which a number certainly bigger than one are instrumental in our organisms’ success in the world. In perception, there is no unique, truth-warranting success.

3.3 No Clear Way from Commonality and Veridicality to Realism

Last for this section, my third counter-argument to Votsis’ design. Again, this argument does not regard O, but aspirations for using his proposed experiment as a ladder towards realism. My arguments in Sects. 3.1 and 3.2 notwithstanding, suppose that a way to convey similarity judgments in under 150 ms is found, and that the experimental results go the way of the objectivist, pinpointing the existence of a common, veridical observation core. What I argue here is that *even this* does not give a clear

advantage to realism. To see why, let us dip a little more into what happens during the first 150 ms of the observation circuit's activation, again without getting too tangled up in the technical details (a thorough account can be found in Raftopoulos, 2019, chapter 5, Sect. 3). The framework below is proffered by Kosslyn (1994) and is also espoused by Raftopoulos (2019, chapter 5), who is, again, on the side of cognitive impenetrability of perception. Within 150 ms after stimulus presentation, low spatial frequency information, semantically processed in the ventral pathway and inferior temporal cortex of the brain, re-enters the extrastriate visual areas, facilitating analysis of high spatial frequency information via specifying certain cues in the image that may facilitate target identification (Barr, 2009; Kihara & Takeda, 2010). Semantic information concerning the putative identity of the surveyed objects enters the observation circuit from this point on. Following, these hypotheses are tested against information from 'low' (non-semantic) circuits, leading eventually to the recognition of objects in the scene (Raftopoulos, 2019, p. 262). Raftopoulos (2019, p. 267) concludes that this synergy between high and low level processes denotes the mark for when cognitive states start to feature in the visual process:

Since the construction of the representations of the putative causes of the perceptual inputs in late vision takes place through the synergy of bottom-up processing transmitting information registered at the lower levels or prediction errors, and top-down processing transmitting information relevant to the testing of hypotheses concerning the probable causes of the input, and in so far as the processes constructing these hypotheses are informed by high-level knowledge about worldly objects, visual perception unifies cognition and late vision; these two become intertwined. (Raftopoulos, 2019, p. 267).

To the extent that this account is correct,¹¹ 150 ms is the final limit for cognitive impenetrability, and the observation core has to be constricted by a timeframe this short if it has any hope of engulfing what we see, and leaving out what we know. People differ cognitively as they endorse different theories proper, and so their observation cores, if defined to be shaped beyond 150 ms, will differ. Despite original experimental plans by Votsis, the contents of the observation core *have* to be redefined as shaped up to 150 ms post-stimulus to have any hope of being theory-free. Here comes, then, the crucial question: how does one make science with contents shaped 150 ms after stimulus onset at maximum? Once one realises how limited such a short time span makes the availability of visual tools, one should be truly at a loss with regard to forming an adequate reply. Recognizing anything as anything is out of the question: non-cognitively penetrated content is just patterns of features, visual images only just not purely kaleidoscopic. Pylyshyn even puts it that perception (maximally what is not cognitively penetrable according to the above model) has, due to the lack of involvement of concepts, non-representational contents; possibly “codes for proximal properties involved in perception, such as edges, gradients, or the sorts of labels that appear in early computational vision” (Pylyshyn, 2007, p. 52). Raftopoulos (2019, p. 94) writes that this paves the way to the view that perceptual states may not have *any* contents, and Fodor and Pylyshyn (2015) argue that perceptual states involve concepts

¹¹ Raftopoulos (2019, p. 265) admits that this account is only in most cases dominant.

only in so far as they refer directly to the world, without any sense or meaning. This issue aside, it is, I think, entirely hopeless to go for a line of science-making without concepts, categories, and even objects. Take any scientific endeavour of choice, as simple or as complicated as one likes, from conjecturing Newton's laws by watching objects move and fall, to inferring the existence of electrons by watching radioactive atoms in a cloud chamber, to testing theoretical predictions in the frame of any of the eight experiments at CERN's Large Hadron Collider. None are remotely conceivable without the involvement of recognizing objects and applying concepts, not to mention reference to unobservable entities. Indeed, as the logical positivists found out quite some time ago, making a case for science-crafting based solely on bare observation is likely an insurmountable challenge. It seems obvious that the difficulty of this project gets multiplied many times over by removing, beyond unobservables and bridge principles, also recognizing, objects, categorization, semantics at large. The bottom-up kind of realism inspired by attempts like Votsis' requires cognitively purified material, but this material is impressively inadequate for literally even the simplest scientific theories. It seems then that we are in a double bind: if the observation core is defined to have contents shaped after 150 ms post-stimulus, then it is cognitively penetrated and theory-laden.¹² If it is defined to have contents shaped until 150 ms after stimulus presentation, it does not even matter whether its contents are cognitively penetrable, theory-free et cetera: these contents are just improper for science. To add to this, the burden of proof lies with the observation core veridicalist, aspiring scientific realist, who has to provide at least a sketch of how one is to get to scientific theory-neutrality from observation core neutrality. This is a quite common admittance, even among veridicalists about perception in the relevant literature (e.g. Raftopoulos, 2008).

At this juncture, I should note that the kind of antirealism hinted at herein does not foremostly regard, as the scientific realism discussion usually does, unobservable entities detected by highly sophisticated experiments. On the contrary, the two most central arguments of this paper that will also feature below—the ushering in of concepts during a very early phase of the observation circuit, and perceptual pluralism—challenge chiefly realism about *macroscopic* objects.¹³ Indeed, if the standard perceptual contents are not uniquely successful and truth-conducive, either due to concepts viciously (for realism) intervening or due to perception having many viable alternatives, it seems that realism about what our percepts correspond to should be affected first and foremost. Naturally, one could here argue that realism about unobservable objects is built on realism about macroscopic objects, and thus traditional scientific realist theses are undermined as well. Certainly, however, related analysis greatly outspans the scope of the present paper, as does the overarching discussion on how the scientific realism debate should be informed by argumentation found herein (some general remarks are, however, presented in the conclusion).

Overall, in this section, I provided three arguments against testing. The first demonstrated that O is true. The second demonstrated that perceptual pluralism undermines

¹² A possible objection against my first counter-argument to Votsis could also surface here, namely that the application of basic perceptual concepts is not ladenness-relevant. The same retorts to this objection, provided in Sects. 3.1 and 3.2, apply here. This also means that some of the argumentative force of my third counter-argument to testing turns on the validity of the first and second ones.

¹³ I thank an anonymous reviewer for pointing this out.

convergence and veridicality regarding the observation core. The third demonstrated that veridicalist-leaning experimental results (in Votsis' sense) show no way forward to scientific realism. The first argument's claim pertinent to perceptual learning was shown to be dependent, to a limited extent regarding the validity of perceptual pluralism, on the validity of the second-counter-argument. The second argument is independent of the other two. As noted in footnote 12, the third counter-argument depends on the first and second ones to the extent that the application of perceptual concepts is relevant to theory-ladenness.

4 Refereeing

Refereeing comes as an empirical suggestion to address D:

D: The contents of observation¹⁴ are a function of already held theories.

This angle does not hold that D is false *sensu stricto*. Refereeing (specific approaches found in e.g. Brewer, 2015; Franklin, 2015 among others) treats theory-ladenness as a kind of *bias*. This bias can be detected in the event of a clash between theories and, more importantly, set aside by an impartial adjudicator. Therefore, even if D occasionally obtains, it does not come with a strong relativist bite. In this section, I argue that refereeing does not work, largely because it misses the depth of theory-ladenness. I provide three counter-arguments to the refereeing approach. The first two follow refereeing in allowing that observation is not necessarily theory-laden,¹⁵ and that ladenness can be detected and removed when present. However, once the in-principle possibility of theory-ladenness has been granted, *detecting* whether theory-ladenness is or is not present, and judging whether the referee or the community of referees carry or do not carry a theoretical bias themselves, are problems with unstraightforward solutions. The third counter-argument introduces what we know from the previous section: perception is decidedly not theory-free post 150 ms after stimulus onset, and this ladenness is carried downstream in the observation circuit. After this short timeframe has elapsed, concepts infiltrate the observation circuit via recognizing and meaning ascription to the scene, and cognition and observation become inextricably intertwined. Thus, even if construed as a bias, theory-ladenness is not a removable one, and treating it as such is an unacceptably superficial approach, missing how deeply concepts cut.

The refereeing approach draws its force from examples within the history of science, which allegedly provide proof of principle that observations can be cleansed from theory. The examples come mostly from clashes in physics, where the deserved winner was decided by producing the right set of observations. The most well-known such case is the 'discovery' of N-rays (Klotz, 1980; Nye, 1980), now considered to be a fictional entity. This supposed new form of radiation was announced in 1903 and, according to the French physicist Blondlot, could penetrate wood and iron but was blocked by water. Brewer (2015, p. 132) notes that over 300 published papers specifying the

¹⁴ Reminder that 'observation' is here meant to signify the third stage in the overall observation process.

¹⁵ This, of course, I grant only for rhetorical purposes, as it directly overrides my argumentation in the previous section.

characteristics of N-rays followed. Unfortunately, soon enough N-rays appeared to be a domestic phenomenon, as they could not be observed outside of France, which development made the American physicist Wood pay a visit to Blondlot's laboratory. There, he modified the observing apparatus in a way that prohibited the production of N-rays (according to the N-ray theory itself). After the French test subjects kept observing indications of the presence of N-rays despite Wood's modifications, Wood published this very experiment as proof that N-rays were the product of the observers' beliefs, not of the physical world's manipulation. The scientific community largely accepted this proof (Brewer, 2015, p. 133), and talk of N-rays was soon to be found in history textbooks and articles like the present. The other well-known instance of putatively successful refereeing regards the dissent between Cambridge and Vienna physicists in the 1920s. Cambridge physicists Rutherford and Chadwick had at the time discovered what we now call artificial nuclear disintegration by noticing that a narrow range of elements emitted protons when bombarded with alpha particles (Stuewer, 1985 for a more detailed account). As Stuewer writes, Rutherford held this to be a hard perceptual task and, in a letter to Bohr, highlighted that "the experiments look easy, when they are really very difficult and full of pitfalls for the inexperienced. So much is this so that I have decided not to get any other work done except under my personal eye" (Stuewer, 1985, p. 255). Simultaneously, in Vienna, Austrian physicists protested that protons were emitted from practically every element upon impact with alpha particles. Rutherford's deciding move was much like Wood's: he travelled to Vienna and conducted the same experiment having removed the source of alpha particles. This newest development remained unbeknown to subjects responsible for counting the scintillations indicating artificial nuclear disintegration, who, unfortunately for the Vienna group, continued to report presence of such scintillations! The case was closed soon after and it was decided that, as Rutherford's account to Chadwick goes, the Vienna observers "saw what they were expected to see" since they knew that certain observations would vindicate the view of their home group (Stuewer, 1985, p. 288).

These and alike incidents (see e.g. Heidelberger, 2003; Franklin, 2015 for more) are often seen as triumphant for making the case for the (potential) objectivity of observation. This is so because they suggest an empirical, concrete way of doing away with theory-ladenness via careful experimental design. By narrating these cases, the anti-relativist no longer needs to implausibly deny the theory-ladenness of observation in principle. They merely have to claim that whatever theory-ladenness there may be can be done away with by bringing an impartial outsider into the laboratory and/or making the observers blind to the desired result of the experiment (Brewer, 2015, p. 134). Indeed, these strategies are not mere hypotheticals; they often consist implemented methodologies of attempting to secure impartiality within science. The double-blind protocol in medical research is a prime example of such a methodology.

I believe that this family of arguments is inadequate for addressing theory-ladenness, this concern fuelled by three sources. First, notice that in the above examples, candidate cases for un-ladenness arise only after the presence of a disagreement between professionals of an esteemed scientific calibre. However, should we follow Kuhn's (1962/1970, chapters III and IV) epistemology of distinguishing between periods of normal science and scientific crisis, we should note that periods of protracted agreement far outspan instances of structural disagreement, which are in historical view

merely moments. That is, due to their common training within the paradigm of their time and familiarization with shared experimental apparatus, scientists tend to agree more often than to disagree on fundamental issues such as what is observed within an experimental instance. Here is then the question: in the absence of disagreement regarding observations, are we to conclude that there is no vicious ladenness to remove, or that there *is* a common vicious ladenness that no one is (yet) able to discern? Since the possibility of theory-ladenness has been granted, the dilemma does not have an obvious answer. What we *do* know from the history of science is that protracted agreements that seem, by our lights, to be agreements on the wrong thing exist. Van Helden (1974) details how, for more than 40 years after Galileo discovered the moons orbiting Jupiter, astronomers kept seeing similar moons around Saturn. It was only after Huygens hypothesized Saturn's ring that astronomers were finally able to see a ring, and not moons, around Saturn. More examples abound (consider also the 300 papers on N-rays mentioned above). It could be only a matter of time, that is, before a new paradigm, currently an unconceived alternative (see Stanford, 2006 for the concept), mandates an organisation of our percepts into observations different than we currently make. In turn, this new set of observations would corroborate theories different than we currently hold, as was the case with Saturn's moons turned ring. Thus, the past cases that now seem like successful cases of theory-ladenness removal cannot warrant that our *current* landscape of scientific theories is not a minefield of theory-ladenness, and thus can neither warrant an anti-ladenness conclusion. Notice that, as I underlined in the beginning of this section, this objection does not turn on any assumptions about the unavoidability of theory-ladenness; it only pinpoints that once the possibility of theory-ladenness is demonstrated, we cannot confidently testify to its having been removed.

My second objection to refereeing goes as follows. Suppose that, as it happened with the N-rays and the Cambridge versus Vienna case, a disagreement from a respectable objector surfaces, and a referee, or better yet a refereeing community, is called to the rescue. Abiding by the logic of the first counter-argument, how are we to conclude that the refereeing community is *not* theory-laden, and that no one is (yet) able to discern this ladenness? Even if we were to somehow determine that we are working with a hundred percent honest referees who reassure us of their impartiality between judged theories, this will not do. As Chadwick reports back to Rutherford regarding the Cambridge-Vienna discord, for the Vienna observers "there was no question of cheating. Rather, they were deluding themselves" (Stuewer, 1985, p. 288). That *consciously* held theories are kept at bay and not coupled with observational results is a necessary but *not* sufficient condition for theory-neutrality. Evidence from cognitive neuroscience suggests that masked, non-consciously registered primes can affect observations (Petit et al., 2006). Recent behavioral, neuropsychological, and neurophysiological data indicates that priming modulates vision at multiple sites along the visual pathways (Kristjánsson & Campana, 2010). Negative affect inhibits semantic and affective priming (Storbeck & Clore, 2008), and priming also affects attentional processes, facilitating target detection and selection on repetition trials (Becker, 2008). The list documenting suchlike effects stretches far further. Here, the objectivist will be quick to note that a majority of science-relevant observations are performed by scientific instruments, not the suggestible human mind, and thus are not susceptible to the

argument I just provided. My reply is that the ladenness of human observations is carried forward to instruments, in a way not unlike that in which ladenness of perception is carried downstream to observation. Due to space constraints, I elaborate this point properly in another work, following others. In a nutshell, the argument is that instruments are initially developed on a basis of measurement-human sense correlation, that this reference is to an important degree retained across a given instrument's life course (see Chang, 2007; Hoel & Carusi, 2017 for related discussions) and also that it is, in the final analysis, also a manifold of human senses that registers, processes, ascribes importance to, instrument measurements (for related remarks see Carusi 2012; for a more comprehensive analysis of this issue and the carrying-forward of ladenness see Pantazakos, 2019b, chapter 3; Vallor, 2009; for an influential overall discussion of the issue see Heidelberg, 2003).

I believe that the above considerations amount to sizeable problems for upholders of refereeing. Even if theory-ladenness is of the ilk assumed by refereeing, i.e. a 'top-layer' bias removable by adjudication, my two arguments above show that knowing when ladenness has been removed and assuring the impartiality of the referees pose considerable challenges. At the same time, an independent, more structural obstacle stands in the way of refereeing, which pertains directly to what was established in the previous section. To bring back to mind what is important here while sparing the reader a repetition of the technical details: perception is cognitively penetrable by information from the higher cognitive centres very early on, the effects of this penetrability are carried downstream to observation, and are not innocuous regarding theory-ladenness. Post-150 ms after stimulus presentation, perception becomes interwoven with cognition, and the effects of cognition are felt all along the subsequent outputs of the visual system. Consider now refereeing, an act quite impossible to carry out within 150 ms or less. The referee charged with the task is supposed to remove any ladenness from the theory or theories they adjudicate by producing their own observation reports. However, to do that, they will have to do things that take far more than 150 ms. To intervene, the referee has to, very minimally, recognize objects proper, ascribe meaning to what they see, so on and so forth. Therefore, they inevitably carry the input of their own cognitive states planted within their own observation reports. Judgments of this kind just cannot be completed in a time frame that renders them theory-free; just like in testing, refereeing judgments, like similarity judgments, are not the kind of judgments that can be meaningfully produced fast enough. Theory-ladenness, that is, is naively treated if treated as some kind of bias that some carry but others do not, a bias that can be cast away by those who do not carry it. Moreover, and if I am right in my argumentation pertaining to perceptual pluralism in the previous section, observation is not ladenness-innocuous even *before* the 150 ms mark. The existence of many successful observation cores, pinpointed by the examples of perceptual variations that do not carry an accompanying disease or pathology, threaten realism by undermining just *which* observation core is appropriate for the summoned referee. Evidently, referees with different observation cores will make different refereeing judgments in different at least some contexts, seeing as observation cores refer to the formation of basic percepts.

At best, ladenness enters the observation circuit after the 150 ms mark. At 'worst', if the observation core has many eligible alternatives, the observation is situated

and laden from the outset. In both cases, ladenness takes hold from the root, and is not a contingent quality of observation that can be removed by a laden-free referee; such treatment of theory-ladenness is unacceptably superficial. The usual objections regarding these arguments could be raised here too: for the best case scenario, plain concepts may not introduce vicious ladenness; for the ‘worst’ case scenario, perceptual pluralism does not provide genuine observation core alternatives because inferior perceptions to the standard are less successful, and superior perceptions produce similar percepts. Counter-arguments to these objections have been provided in Sects. 3.1 and 3.2 of the present paper. To the extent that these obtain, so does my third counter-argument to refereeing.

5 Conclusion

In the course of the present paper, I interrogated two recent versions of empirical methods aiming to solve the epistemological problem that certain strands of theory-ladenness pose. Method one, testing, addresses the ladenness of observation core by expertise. It contends that, given favourable experimental results, one may ascertain that the contents of the observation core converge between laymen and experts, and that they are veridical. Testing does not work for three reasons. First, research in cognitive neuroscience demonstrates that expertise infiltrates the observation process from very early on; perceptual learning affects perception from the onset, while adoption of theories proper affects perception post-150 ms after stimulus onset. Since, according to Votsis’ design, the experimental tasks take way more than 150 ms to be completed, and subjects with very different perceptual trainings partake of the experiment, the observation core will turn out to be expertise-laden. Second, perceptual pluralism ensures that there is no convergence across the contents of observation cores of subjects in different perceptual states. Moreover, it is implausible that all observation cores except one can be epistemically discredited as less successful and truth-warranting, so as to argue that convergence within typical perception is a solid foundation for its veridicality. Third, even if convergence of the observation core’s contents is somehow ascertained, and even if these contents are veridical, venturing from this veridicality to scientific realism is a way undiscovered, if not inexistent. The last argument does not undermine testing per se, but rather testing as a step within a scientific realist research programme.

Method two, refereeing, addresses the ladenness of observation by already held theories. Refereeing does not deny the existence of such ladenness, but suggests that it can be detected and corrected for by virtue of a refereeing agent. Refereeing does not work either, for three reasons. First, even granting that there is a non-theory-laden way of observing, contrary to my argumentation in testing, the how of ascertaining that all related ladenness has been removed is unclear. Second, the how of determining the impartiality of the referee is also unclear. Third, and most importantly, theory-ladenness comes not only in the form of some kind of bias and/or conscious theoretical disposition towards a certain observation. It occurs, as mentioned many times over by now, fast enough in the observation circuit to forbid hopes of a non-laden referee

coming in and doing ladenness-cleaning work. Any referee summoned to the job will inevitably their own ladenness while refereeing.

As I presented them here, both approaches commit two central mistakes. One is to underestimate how promptly in the observation circuit theory-ladenness comes in. To amend this mistake, future literature about theory-ladenness is suggested to account for cognitive neuroscientific work on the issue, currently documenting clearly that theory-ladenness creeps in our observation circuit from very early on, possibly early enough to undercut any heavy-duty realist claims about science. The second mistake is to sidestep the plurality of modes of perceiving the world and the success involved therein. It is my view that future approaches to theory-ladenness, objectivity, scientific realism, and their interconnections, should account foremostly about these multifarious pluralisms that have been highlighted as irremovable parts, even assets, of the scientific endeavour. I am eager to see what prospects may arise for realism after its being pluralistically (and, on a closely related note, perspectivally) informed, with suchlike projects being well underway (e.g. Chang, 2017; Giere, 2010; Glick, 2019; Massimi, 2019).

Funding Not applicable.

Availability of data and material Not applicable.

Declarations

Conflicts of interest Not applicable.

Code availability Not applicable.

Consent to participate Not applicable.

Consent for publication Given by the author submitting the paper to the journal, no other authors involved.

Ethics approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barr, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society London b: Biological Sciences*, 364, 1235–1243.
- Becker, S. (2008). The mechanism of priming: Episodic retrieval or priming of pop-out? *Acta Psychologica*, 127(2), 324–339.

- Block, N. (2013). The grain of vision and the grain of attention. *Thought, 1*, 170–184.
- Blom, J. D. (2010). *A dictionary of hallucinations*. Springer.
- Bogen, J. (2009). Theory and observation in science. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2020 Edition). <https://plato.stanford.edu/archives/sum2020/entries/science-theory-observation/>.
- Brewer, W. (2015). Perception is theory laden: The naturalized evidence and philosophical implications. *Journal for General Philosophy of Science, 46*, 121–138.
- Carusi, A. (2012). Making the visual visible in philosophy of science. *Spontaneous Generations: A Journal for the History and Philosophy of Science, 6*(1), 106–114.
- Chang, H. (2007). Scientific progress: Beyond foundationalism and coherentism. *Royal Institute of Philosophy Supplement, 61*, 1–20.
- Chang, H. (2012). Acidity: The persistence of the everyday in the scientific. *Philosophy of Science, 79*(5), 690–700.
- Chang, H. (2017). Is pluralism compatible with scientific realism? In J. Haatsi (Ed.), *The Routledge handbook of scientific realism*. Routledge.
- Churchland, P. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science, 55*, 167–187.
- Ffytche, D., Blom, J. D., & Catani, M. (2010). Disorders of visual perception. *Journal of Neurology, Neurosurgery, and Psychiatry, 81*, 1280–1287.
- Fodor, J. (1984). Observation reconsidered. *Philosophy of Science, 51*, 23–43.
- Fodor, J. (1988). A reply to Churchland's perceptual plasticity and theoretical neutrality. *Philosophy of Science, 55*, 188–198.
- Fodor, J., & Pylyshyn, Z. (2015). *Minds without meanings: An essay on the content of concepts*. MIT Press.
- Franklin, A. (2015). The theory-ladenness of experiment. *Journal for General Philosophy of Science, 46*, 155–166.
- Giere, R. N. (2010). *Scientific perspectivism*. University of Chicago Press.
- Glick, D. (2019). Pluralist structural realism: The best of both worlds? *Synthese, 198*, 4145–4166.
- Gould, S. J. (1981). *The mismeasure of man*. W.N. Norton and Company.
- Grandin, T. (2012). *Different, not less: Ultimate success stories from people with autism and Asperger's*. Future Horizons.
- Happé, F., & Frith, U. (2006). The weak coherence account: Detailed-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25.
- Heidelberger, M. (2003). Theory-ladenness and scientific instruments in experimentation. In H. Radder (Ed.), *The philosophy of scientific experimentation*. University of Pittsburgh Press.
- Higashida, N. (2013). *The reason I jump*. Penguin Random House.
- Hoel, A. S., & Carusi, A. (2017). Merleau-Ponty and the measuring body. *Theory, Culture, & Society, 35*(1), 45–70.
- Kihara, K., & Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision Research, 50*, 2158–2162.
- Klotz, I. M. (1980). The N-ray affair. *Scientific American, 242*(168–70), 173–175.
- Kosslyn, S. M. (1994). *Image and brain*. MIT Press.
- Kras, J. (2010). The “ransom notes” affair: When the neurodiversity movement came of age. *Disability Studies Quarterly, 30*(1).
- Kristjánsson, A., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics, 72*, 5–18.
- Kuhn, T. (1962/1970). *The structure of scientific revolutions*. The University of Chicago Press.
- Martin, C. C., & Gordon, R. (2010). The evolution of perception. *Cybernetics and Systems, 32*(3–4), 393–409.
- Massimi, M. (2019). Realism, perspectivism, and disagreement in science. *Synthese*. <https://doi.org/10.1007/s11229-019-02500-6>
- Mottron, L., Dawson, M., & Soulières, I. (2009). Enhanced perception in savant syndrome: Patterns, structure and creativity. *Philosophical Transactions of the Royal Society b: Biological Sciences, 364*(1522), 1385–1391.
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*(1), 27–43.

- Nye, M. J. (1980). N-rays: An episode in the history and psychology of science. *Historical Studies in the Physical Sciences*, 11, 125–156.
- Pantazakos, T. (2019b). *Scientific realism and phenomenology through the case study of autism* (Doctoral Dissertation, UCL (University College London)).
- Pantazakos, T. (2019a). Treatment for whom? Towards a phenomenological resolution of controversy within autism treatment. *Studies in History and Philosophy of Science Part c: Studies in History and Philosophy of the Biological and Biomedical Sciences*, 77, 101176.
- Petit, J.-P., Midgley, K., Holcomb, P. J., & Grainger, J. (2006). On the time course of letter perception: A masked priming ERP investigation. *Psychonomic Bulletin & Review*, 13, 674–681.
- Pylyshyn, Z. (2007). *Things and places: How the mind connects with the world*. MIT Press.
- Quartz, S., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20(4), 537–596.
- Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cognitive Science*, 25, 423–451.
- Raftopoulos, A. (2008). Perceptual systems and realism. *Synthese*, 164(1), 61–91.
- Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press.
- Raftopoulos, A. (2015). The cognitive impenetrability of perception and theory-ladenness. *Journal for General Philosophy of Science*, 46, 87–103.
- Raftopoulos, A. (2019). *Cognitive penetrability and the epistemic role of perception*. Palgrave Macmillan.
- Ripamonti, L. (2016). Disability, diversity, and autism: Philosophical perspectives on health. *The New Bioethics*, 22(1), 56–70.
- Ropar, D., & Mitchell, P. (1999). Are individuals with autism and Asperger's syndrome susceptible to visual illusions? *Journal of Child Psychiatry and Psychology*, 40, 1283–1293.
- Sacks, O. (1985/2011). *The man who mistook his wife for a hat*. London: Picador.
- Sacks, O. (1995). *An anthropologist on Mars*. Knopf.
- Schurz, G. (2015). Ostensive learnability as a test criterion for theory-neutral observation concepts. *Journal for General Philosophy of Science*, 46(1), 139–153.
- Siegel, S. (2011). Cognitive penetrability and perceptual justification. *Nous*, 46(2), 1–22.
- Stanford Neurodiversity Project. (n.d.). *About the Stanford Neurodiversity project*. Retrieved February 8, 2021, from <https://med.stanford.edu/neurodiversity.html>.
- Stanford, K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. Oxford University Press.
- Storbeck, J., & Clore, G. L. (2008). The affective regulation of cognitive priming. *Emotion*, 8(2), 208–215.
- Stuewer, R. H. (1985). Artificial disintegration and the Cambridge–Vienna controversy. In P. Achinstein & O. Hannaway (Eds.), *Observation, experiment, and hypothesis in modern physical science*. MIT Press.
- Vallor, S. (2009). The fantasy of third-person science: Phenomenology, ontology and evidence. *Phenomenology and the Cognitive Sciences*, 8, 1–15.
- Van Helden, A. (1974). Saturn and his Anses. *Journal for the History of Astronomy*, 5, 105–121.
- Votsis, I. (2015). Perception and observation unladen. *Philosophical Studies*, 172, 563–585.
- Votsis, I. (2020). Theory-ladenness: Testing the 'untestable.' *Synthese*, 197, 1447–1465.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.