European Journal for Philosophy of Science (2023) 13:43
https://doi.org/10.1007/s13194-023-00544-7

PAPER IN THE PHILOSOPHY OF THE SCIENCES OF MIND AND BRAIN

Afactivism about understanding cognition

Samuel D. Taylor[1]

Received: 16 January 2023 / Accepted: 19 July 2023 / Published online: 6 September 2023

Abstract
Here, I take alethic views of understanding to be all views that hold that whether an explanation is true or false matters for whether that explanation provides understanding. I then argue that there is (as yet) no naturalistic defence of alethic views of understanding in cognitive science, because there is no agreement about the correct descriptions of the content of cognitive scientific explanations. I use this claim to argue for the provisional acceptance of *afactivism* in cognitive science, which is the view that the truth or falsity of an explanation of cognition is irrelevant to whether that explanation provides understanding. I conclude by discussing the relation between understanding in cognitive science and understanding in other domains.

Keywords Cognitive science · Understanding · Afactivism · Alethic · (Non-)Factivism

1 Introduction

I take alethic views of understanding to be all views that hold that whether an explanation is true or false matters for whether that explanation provides understanding. Such alethic views are popular—if not, entirely dominant—in the contemporary debate about the understanding we gain from scientific explanations.[1] Still, there is disagreement about the relationship between the truth/falsity of an explanation and that

[1] Here, I am concerned with scientific explanations in particular, but I will often use the term 'explanation' below for convenience.

✉ Samuel D. Taylor
samuel.da.taylor@gmail.com

[1] Department of Philosophy, University of Kent, Cornwallis North West, CT2 7NF Canterbury, UK

explanation's capacity to provide understanding. More precisely, there is ongoing disagreement about whether an explanation *must* be true in order for that explanation to provide understanding.

In this paper, I argue that the contemporary debate about the relation between the truth of an explanation and the understanding provided by that explanation misses the point in the context of cognitive science, because there is (as yet) no naturalistic defence of alethic views of understanding in this domain. This is the case, I claim, because there is no agreement about the correct descriptions of the true—and, hence, not idealized—content of cognitive scientific explanations. I use this claim to argue for a provisional acceptance of *afactivism* in cognitive science, which is the view that the truth or falsity of an explanation of cognition is irrelevant to whether that explanation provides understanding. I conclude by discussing the relation between understanding in cognitive science and understanding in other domains.

2 Alethic views of understanding: strict-factivism, non-factivism, and quasi-factivism

In the present debate, there are three, alethic views of understanding available: factivism, non-factivism, and quasi-factivism. Factivists argue that we only arrive at scientific understanding when we "grasp" a *correct* explanation (de Regt, 2009b; Strevens, 2011; Trout, 2002). In this vein, Trout (2007, 585-586) claims that "scientific understanding is the state produced, and only produced, by grasping a true explanation." Likewise, Strevens (2013, 1) argues that "An individual has scientific understanding of a phenomenon just in case they grasp a correct scientific explanation of that phenomenon."

In contrast to factivists, non-factivists argue that we can arrive at scientific understanding when we "grasp" an explanation that is, strictly speaking, incorrect (Elgin, 2007, 2009). On this view, scientific understanding and truth come apart, because scientific understanding can be provided by false explanations. A further view—quasi-factivism—holds that the scientific understanding provided by an explanation is tied to truth, but that this need not be the truth of the relevant explanation itself, because the truth of judgements about the explanation are also relevant.

To get to grips with the factivist, non-factivist, and quasi-factivist positions, consider the following proposition:

(1) Scientific explanation p provides scientific understanding of x.

Where p is the base of understanding (the thing providing understanding) and x is the object of understanding (the thing understood) (Rancourt, 2017).[2] The first thing to note is that all participants in the debate about the facticity of scientific

[2] Note that the basis of understanding can be, but need not be, an explanation; and the object of understanding can be, but need not be, an explanandum. Doyle et al. (2019), for example, argue that "the ideal gas law is best interpreted as the object of understanding, with statistical mechanics serving as its proper basis."

understanding would (likely) assent to the claim that there are true explanations that provide understanding.[3] Thus, they will all at least assent to the following claim:

$$\exists p \exists x (Tp \wedge U(x, p)) \tag{a}$$

Where T is the predicate 'is true' and $U(x, p)$ is the binary predicate 'scientific explanation p delivers scientific understanding of x.'

According to "strict-factivism," however, the base of understanding—e.g. the scientific explanation p—must be true in order for (1) to be true, and this holds whether we take p to be a single proposition or to be a conjunction of propositions (Doyle et al., 2019). So, for the strict-factivist:

$$\forall p \exists x (Tp \leftrightarrow U(x, p) \tag{b}$$

According to Rice (2016, 85, my italics):

> Among philosophers of science, it is widely accepted that understanding is factive in the sense that (*at least some of*) the beliefs (or propositions) within one's understanding must be true (Grimm, 2006, 2012; Mizrahi, 2012; Strevens, 2011, 2013).

This reference to "at least some" leaves open the question of how we are to differentiate factivism from non-factivism, where the latter holds that there are at least some cases where (1) is true and p is false.[4] If p is a single proposition, then we can easily make sense of the difference between factivism and non-factivism. But if we think of p as a conjunction of propositions, then the difference is obscured, since p will be false if only one of its conjunct is false, even if it remains the case that *at least some of* its conjuncts are true. In any case, non-factivists defend a claim that strict-factivists do not; namely:

$$\exists p \exists x (\sim Tp \wedge U(x, p)) \tag{c}$$

Additionally, the non-factivist will assume what Doyle et al. (2019, 346) call the "Parity Condition":

> The understanding of x resulting from either not accepting p or accepting a more accurate proposition instead of p is not better than the understanding provided by accepting p.

The central idea here is that "p provides understanding of x if and only if understanding of x results from accepting that p." So (c) can be modified to:

$$\exists p \exists x ((\sim Tp \wedge Ap) \wedge U(x, p)) \tag{c'}$$

[3] I say 'likely' here, because, as we will see below, it is possible, but highly unlikely, that non-factivists could reject this clam.

[4] It is possible, but highly unlikely, that the non-factivist could hold that in all cases in which (1) is true p is false, which would be to reject (a).

Where A is the predicate 'is accepted.'

"Quasi-factivists" reject this "Parity Condition" and hold instead that in our evaluation of the truth of (1) we should not conflate "understanding a model with understanding the phenomena represented by a model," because "a true proposition describing a more complex relationship between [the model] and reality [can be] the authentic basis of understanding" (Doyle et al., 2019, 348) (see Greco, 2013; Kvanvig, 2009; Mizrahi, 2012). We can call this putatively "authentic basis" p^* and think of it as a judgement about the explanation/model in question. Thus, quasi-factivists will sometimes appeal to second-order propositions—e.g. p^*—to function as the basis of understanding and account for how (1) can be true even when p is false and not (entirely) accepted.

The idea is that p^* will be a second-order proposition—e.g. that 'p is approximately true' or that 'p gives us a better grasp of the correct explanation'—, which provides better understanding in virtue of being more accurate than p. To make sense of this, consider the following example from Greco (2013, 297):

> let p be that it is 3 o'clock. We are happy to say, 'S knows that p,' even when p is strictly false, because it is in fact 3:01. Why? Because p is 'true enough.'

Here, the quasi-factivist argues that even though p is false and so is not (entirely) accepted, a second-order proposition about p—namely p^*: 'p is true enough'—is true and so is accepted. The idea of quasi-factivism, then, is that when we say that a false scientific explanation, p, provides understanding of x, we really mean that we gain understanding of x from some authentic basis, p^*, which makes a true judgement about p (e.g. that p is true enough). Thus, the quasi-factivist claims that in every case where (1) is true either:

$$\exists p \exists x (Tp \leftrightarrow U(x, p)) \tag{d}$$

Or:

$$\exists p \exists p^* \exists x ((\sim Tp \wedge \sim Ap) \wedge (Tp^* \wedge Ap^*) \wedge U(x, p^*)) \tag{e}$$

3 Scientific understanding and idealization

There are, then, key differences between the strict-factivist, non-factivist, and quasi-factivist positions. Still, all accept that most–if not all—scientific explanations do not represent the world as it really is, because they employ *idealizations* that are, strictly speaking, false (Odenbaugh, 2011, 1187). As Rice (2016, 82-83) explains:

> idealizations are known to be false assumptions—they deliberately misrepresent or distort the features of real-world systems. [...] The general idea is that, given that our best models and theories contain known to be false assumptions, even if they make accurate predictions we have little reason to believe they are true.

According to Weisberg (2007, 639-646), idealization involves "the intentional introduction of distortion into scientific theories." The aim of idealization might be to simplify theories in order to make them computationally tractable (so-called Galilean

idealization), to facilitate the construction of theoretical models that include only the core causal factors which give rise to a phenomenon (so-called Minimalist idealization), or to facilitate the construction of "multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon" (so-called Multiple Models Idealization).

In the literature on scientific understanding and idealization, the most discussed example concerns scientists' false assumption that systems consist of a number of identical non-interacting particles in certain derivations of the ideal gas law in statistical mechanics (cf. Doyle et al., 2019, for more detail). But idealizations abound in science. For example, there are idealizations that appeal to entities that do not exists (such as fully rational agents in economics or frictionless planes in physics) or idealizations that subtract important worldly elements from their accounts—e.g., long range intermolecular forces in thermodynamics.

It is helpful to consider how each of the views on understanding introduced in the previous section respond to the wealth of evidence for idealization in science. Consider strict-factivism to begin with. At first blush, strict-factivism may appear to be a non-starter, because we would then have to accept that scientific understanding is scarce or non-existent, because most—if not all—explanations are, strictly speaking, false. However, some have argued that this "idealization argument" against strict-factivism can be diffused if we accept that idealizations "permit the recovery of true content even if an agent is unaware of the idealized nature of the model or theory" (Ross, 2023, 761).

The key idea here is that "understanding is an epistemic state that requires a subject" and that idealizations are useful and convenient "tools for eliciting true beliefs that facilitate understanding of their objects" (Ross, 2023, 766). One can then argue that strict-factivism is not undermined by idealization in science, because "only the truths we extract from [scientific explanations] are elements of the content of our understanding" (Lawler, 2021, 6859). For example, that in optimality models in biology it is only the extracted true beliefs—e.g. 'Natural selection will tend to favour crows which optimise calorific gains when foraging over those which are profligate with their energy—and not the extracted false beliefs—e.g. 'All crow flight is horizontal'—that engender understanding.

Non-factivist, on the other hand, take the presence of idealizations as confirmatory of their view (see Potochnik, 2017). For instance, Doyle et al. (2019, 352) claim to show that some idealized—and so false—scientific models provide us with scientific understanding, because they provide us with four cognitive goods: simpler calculation, highlighting irrelevancies, explanation, and constructing new models. The upshot, then, is that non-factivism is both motivated and plausibly substantiated by the presence of idealization in science.

Quasi-factivists demur. They try to circumvent the problem of idealization by shifting the attention away from idealized—and so false—explanations and towards true judgements about those idealized explanations. Strevens (2013, 512), for instance, argues that:

> idealizations that appear to make some false assumption about the world—for example, the assumption that there are no long-range intermolecular forces (in the ideal gas model) [...]—in fact make true claims about difference-making. When used to explain the approximate truth of Boyle's law, for example, the ideal gas model, when it sets long-range intermolecular forces to zero, is in fact saying, when properly interpreted, that long-range intermolecular forces make no difference to the law's approximately holding.

Strevens' (2017, 42) view is that "Idealizations help us to understand a phenomenon by giving us a better grasp of a correct explanation of the phenomenon." More precisely, Strevens claims that idealizations help us to truthfully model those factors—and only those factors—that make a difference to bringing about the explanandum (e.g. gas expansion in the case of the ideal gas law). Thus, for Strevens, it is our true judgements about the truth-tracking features of idealizations that provide the authentic base of understanding (e.g. p^*) and not the idealized—and so false—explanations (e.g. p) themselves.

4 Describing science

My aim here is not to resolve the debate between strict-, non-, and quasi-factivists about the implications of idealization in science. Rather, I want to put forward an alternative position altogether (in the context of cognitive science): afactivism. However, to justify this move it is necessary to first defend a specific claim; namely, that given the presence of idealization in science, we will only be able to make progress in the debate between strict-, non-, and quasi-factivists when we agree on the descriptions of the content of the relevant scientific explanations.[5]

To describe the content of a scientific explanation is to give some account of the nature and structure of the *content* of that explanation. For example, one might argue, in accordance with the Deductive-Nomological Model, that explanatory content consists of two "constituents": an explanandum "describing the phenomenon to be explained"; and an explanans, which is "the class of those sentences which are adduced to account for the phenomenon" (Hempel & Oppenheim, 1948, 137). Alternatively, one might argue, in accordance with the Causal Mechanical model, that explanatory content consists of (bodies of) information that trace(s) the causal processes and interactions leading up to the event to be explained (see Salmon, 1984).

These coarse-grained descriptions reflect our attempts to describe explanatory content in a way that abstracts away from the contextual features of explainers and their audiences. When I speak of descriptions of explanatory content hereafter, however, I will be referring to something more specific: our specification of some feature of the explanatory content—be that a (set of) proposition(s) or body of information—as an

[5] I will focus here on scientific *explanations* for convenience only and I do not deny that there are a range of other scientific products that could also be described. For example, laws and principles, classification schemes, methods, practices, measures, devices and materials, model organisms etc.

idealization or not. I do not want to take a stand on whether these features are part of a "structural" description of explanatory content or if they are contextual (and it does not matter for my argument below). But I do want to say that our descriptions of these features are crucially important for the debate about facticity of scientific understanding.

There are compelling reasons for thinking that this claim must be true. The first is that the debate between strict-, non-, and quasi-factivists is about the understanding provided by exactly those things—e.g. idealizations etc.—that form part of the explanatory content of scientific explanations. It seems evident, therefore, that to make progress in a debate about the facticity of the explanatory results of science we must be able to agree on the description of the idealized content of scientific explanations. If this were not true, then it would not be clear that we are discussing the same understanding-providing thing.

The second—and more important—reason is that the descriptions of the idealized content of scientific explanations really matter. To make this point concrete, consider a scientific explanation p, whose content can be given description a or description b, where $a \neq b$ and both a and b take a view on the idealizations undertaken in p. Furthermore, suppose that we cannot agree about whether to endorse p_a or p_b, where both take different views on the idealizations included in explanatory content of p. The problem, then, is that it is not clear that we mean the same thing when we say that p provides understanding, because we cannot be sure if we are referring to p_a or p_b respectively.

This point brings into focus the problem with recent attempts to defend strict-factivism by arguing that idealizations are useful and convenient tools for extracting true beliefs, and that only true beliefs engender understanding (see Lawler, 2021; Ross, 2023). This view implies that we can only make sense of what is understood in any instance—i.e. the content of understanding—if we can separate out the true and false beliefs that have been extracted from the relevant explanation. But then it follows that insofar as we cannot agree about whether to endorse p_a or p_b, we will be unable to agree about what is, in fact, understood.

To be clear: it will not help here to say simply that "true [extracted] beliefs that facilitate understanding of their objects," because we will not be able to agree on what the objects are in virtue on not being able to agree on what the extracted true beliefs *are*. For illustration, suppose that we cannot agree about whether the extracted belief that 'All crow flight is horizontal' is (p_a) or is not (p_b) part of the idealized content of the optimality model proposed by Zach (1978). It follows that we cannot agree about whether this belief is true or false. According to "extractive" strict-factivism, therefore, we would not be able to say whether the object understood is, for instance, <crows that drop whelks and optimise calorific gains>, or <crows that drop whelks, optimise calorific gains, *and only fly horizontally*>.[6]

In this way, the presence of disagreements about how to describe the idealized content of scientific explanations forces the extractive strict-factivist to accept that, in those cases where we disagree, we do not know what we understand even when we can assume that we do, in fact, understanding something (since at least some of the

[6] Here, I use <> to denote objects for convenience.

extracted beliefs are plausibly true). Thus, if we endorse extractive strict-factivism, then the presence of disagreements about how to describe the idealized content of scientific explanations risks making many instances of understanding mysterious, because then extractive strict factivism amounts nothing more than a theory of understanding in abstraction, clinging to the notion of truth, but silent on the non-alethic details about what is actually understood.

And disagreements about how to describe the idealized content of scientific explanations also present a problem for non-factivist and quasi-factivist views of understanding. This is the case because we need to decide between p_a or p_b to identify which features of p are idealized. And without identifying which features of p are idealized, we will not be able to determine, for example, whether it really is the case that "Idealizations help us to understand a phenomenon by giving us a better grasp of a correct explanation of the phenomenon."

Consider as another toy example the ideal gas law. The assumption in the literature has been that all can agree on the correct description of the idealized content of this explanation and that the debate between strict-, non-, and quasi-factivists is simply about whether a particular idealization—e.g. that *particles in an ideal gas do not interact*—provides understanding. Strict-factivists hold that the idealization does not provide understanding, because only extracted true beliefs provide understanding. Non-factivists hold that this "strictly false proposition may provide the greatest understanding of ideal gases." And quasi-factivists hold that:

> it is not the falsehood, particles in an ideal gas do not interact, that provides understanding. Rather, a true proposition describing a more complex relationship between this proposition and reality is the authentic basis of understanding the ideal gas law (Doyle et al., 2019, 348-349).

But suppose, for illustration, that we could *not* agree on the correct description of the idealized content of this explanation, and that some held that the proposition, *particles in an ideal gas do not interact* is an idealization (p_a) and others held that it is not (p_b). In this scenario, there would be no way to establish either non-factivism or quasi-factivism without first taking a (disputed) stance on the correct description (e.g. without making a choice between p_a or p_b), because establishing (and refuting) either position requires first that we take a stance on the nature of the proposition(s) providing understanding.

5 Naturalism and normativity

To an extent, this kind of worry has been pre-empted by Doyle et al. (2019, 345-346), who complain that debates about the facticity of scientific understanding:

> have suffered from confusions about the relevant science, as well as from conceptual confusions [...] the proper corrective [for which] is both naturalistic and normative: better descriptions of the science, and more careful analytical work.

Still, their working assumption is that "careful analytical work" will play an equally important role when it comes to deciding between strict-, non-, and quasi-factivism.

Now, one may argue that there is no way to arrive at descriptions of the content of scientific explanations without also engaging in normative analysis of some kind. But even if this is true, we must ensure that this kind of normative analysis is clearly divorced from the normative analysis that aims to defend strict-, non-, or quasi-factivism. If we do not keep these two kinds of normative analysis separate, then we cannot be sure that our descriptions of explanatory content are not tailored to positions in the facticity debate. For example, descriptions that countenance some putative, understanding-providing idealization and so support non-factivism.

I do not mean to deny, therefore, that normative analysis can be of value when we are coming to agree on the descriptions of explanatory content. For example, in the case of the ideal gas law, where all have come to agree that the proposition, *particles in an ideal gas do not interact*, is an idealization. I do, however, want to argue that the success of normative analysis that aims to defend strict-, non-, or quasi-factivism will depend upon us *first* coming to agree on the correct descriptions of explanatory content, because only then can we be sure that all can accept the conclusions of the facticity-relevant normative analysis in question.

The point here is that the relevant normative analysis in the facticity debate will only make sense—and be compelling—against the backdrop of a consensus about the descriptions of the relevant explanatory content. In terms of the example above, it would make little sense to engage in normative analysis to defend strict-, non-, or quasi-factivism while there remains a tension between p_a and p_b, because this normative analysis will only be convincing to all when tied back to a description of the relevant explanatory content that all can accept. Without this agreement, the basis of any normative analysis that defends strict-, non-, or quasi-factivism will be disputed and, hence, the work itself unconvincing.

Note, that I am *not* arguing that we should abandon philosophy of science altogether. Rather, I am arguing that any normative analysis defending strict-, non-, or quasi-factivism must proceed from an uncontested normative analysis that specifies the relevant descriptions of the explanatory content of scientific explanations. Thus, we have two, separate kinds of analysis aligned to two, separate kinds of debates: normative analysis aligned to debates about understanding and normative analysis aligned to debates about how to describe the content of scientific explanations. My claim is that we cannot hope to successfully (and impartially) carry-out the former until we have agreement about the latter.[7]

One may, of course, dispute this point, but then one is forced to admit that it is perfectly acceptable for our theories of scientific understanding to contaminate our descriptions of the content of scientific explanations. This, it seems clear to me, is to challenge the objectivity of science. More specifically, it is to hold that philosophers

[7] There is a further question as to whether both kinds of analysis count as instances of philosophy of science (although the latter surely does), but this is merely a semantic issue that has little, if any, bearing on my argument.

of science are within their rights to interpret the products of science in such a way that coheres with their preferred—and, hence, subjective—account of how those products provide us with understanding. Such a move strips science of its capacity to function as in independent arbiter in the dispute and so should be rejected out of hand.

Thus, I contend that all must accept that progress in the debate between strict-, non-, and quasi-factivists depends upon first coming to an agreement about descriptions of idealized explanatory content, whether this involves normative analysis divorced from the facticity debate or not. I call this position *conscientious naturalism*.[8] This kind of conscientious naturalism might well already be at play in the discussion of the ideal gas law, because there is no disagreement there about the description of the relevant explanatory content. The problem, however, is that not all cases are so easy.

6 Idealization in cognitive science

Those writing on scientific understanding and idealization have focused on a number of examples from science. For instance, the ideal gas law from statistical mechanics. Notably, however, no one in the debate has yet engaged closely with examples from cognitive science, which is surprising since all seem to agree that understanding is a *cognitive* achievement. As Elgin (2009, 35, my italics) puts it "Understanding, then, is in the first instance a *cognitive relation* to comprehensive, coherent sets of *cognitive commitments*."[9]

There is, perhaps, a reason that no one in the debate has engaged with examples from cognitive science: because there is little agreement within cognitive science about where idealization has occurred. Consider the case of mental representations, which are ubiquitous in (folk-)psychology.[10] Coelho Mollo (2020) argues that "cognitive representations are part of an idealized explanatory model." This view

[8] This term was first coined by Doyle et al. (2019), but my usage is somewhat different to—and more demanding than—theirs.

[9] Elgin defends non-factivism, but quasi-factivists like Strevens (2013) also "analyze understanding [...] as a certain kind of (externally valid) mental state."

[10] A representation is commonly taken to be an object with semantic properties (content, reference, truth-conditions, truth-value, etc.), which is just to say that the representation is *about* something other than itself and that the representation can more accurately/truthfully etc. *stand in for* that thing. A number of objects plausibly fall into this category, including some of the products of science, such as models, theories, equations, and even, say, X-ray pictures.
Here, I am referring to a subset of all plausible representations: *mental* representations. There is a long standing debate about the structure or format of mental representations (cf. Machery, 2009; Prinz, 2002; Taylor & Vosgerau, 2021), but, for our purposes, it suffices to think of mental representations as mental objects with semantic properties. Such mental objects—which may include thoughts, concepts, ideas, rules, schemas, or images—are objects that are constitutive of cognitive or mental systems; e.g. brains/minds.
These mental representations are said to be retrieved and used by brains/minds when they undertake cognitive tasks such as, say, categorisation, inductive and deductive reasoning, and perception. With this said, it may be helpful to note that some of the products of cognitive science—e.g. theories, models—may be representations that are (allegedly) *about* mental representations, because these representations are *about* the brain/mind.

is consistent with the position defended by Egan (2014) and would, perhaps, be accepted by non-representationalists such as Hutto and Myin (2017). However, robust representationalists—such as Nanay (2019) and Shea (2018)—would deny that representations are mere idealizations.

Still, one might argue that little agreement does not equate to *no* agreement. But while it is true that there is limited agreement about the presence of idealization in cognitive science, there is no agreement about the *scope* of idealization in cognitive science. To see why, consider the case of Bayesian cognitive science (see Chater & Oaksford, 2008; Clark, 2013a; Colombo & Hartmann, 2017; Taylor & Sutton, 2021). According to Chater et al. (2011, 194):

> The essence of Bayes is the commitment to representing degrees of belief with the calculus of probability. By adopting appropriate representations of a problem in terms of random variables and probabilistic dependencies between them, probability theory and its decision-theoretic extensions offer a unifying framework for understanding all aspects of cognition that can be properly understood as inference under uncertainty: perception, learning, reasoning, language comprehension and production, social cognition, action planning and motor control, as well as innumerable real-world tasks that require the integration of these capacities.

Now, most would accept that the application of Bayes rule in cognitive science is an idealization, because "the Bayesian analysis abstracts away from mechanism—which presumably differs in detail between verbal reasoning, perception, and human and animal learning" (Chater et al., 2011, 196). However, it would be much too quick to say that we therefore have agreement about the scope of idealization in these cases, because all Bayesian approaches posit representations (of degrees of belief). Are these representations idealizations as well? The answer to this question is unclear. So while all may agree that there is some idealization in the case of Bayesian cognitive science, there is (as yet) no agreement about the scope of idealization.

And this concern is ubiquitous in cognitive science. Consider explanations in (cognitive) neuroscience. For example, explanations of image compression in the retina involving first "photoreceptors" and then, after multiple processing stages, retinal ganglion cells (RGCs). According to Thomson and Piccinini (2018, 200):

> While neuroscientists rarely say that individual photoreceptors represent the visual world in primary research publications (Baylor, 1987; Korenbrot, 2012), once we reach the RGCs, such attributions are pervasive (Li et al., 2014; Roska & Werblin, 2001; Soo et al., 2011; Wandell, 1995). That is, once we reach neurons with receptive fields that compactly encode information about a relatively high-capacity stimulus space, the language of representation is used frequently.

While Thomson and Piccinini are adamant that such neural representations are not merely idealizations, recent debates in the literature show that this issue is far from

settled (see Egan, 2014). So, again, we find that there is no consensus about the scope of idealization in cognitive science.

It is important to briefly address one concern here: that my analysis of cognitive science is too superficial and that a more extensive discussion is required to substantiate the conclusion that there is little agreement about where idealization has occurred in cognitive science. Rebuttals of this kind are cheap and commonplace in the literature. But the concern carries little weight here, because an extensive discussion of the intricacies of various cognitive scientific explanations would be superfluous. The point can be made in only a few words: talk of mental representations abounds in cognitive science (see Fusco 2023; Ishikawa and Senju 2023; Li et al. 2023; Yu et al. 2023, as only a few, recent examples) and yet there is no consensus about if—and, if so, to what extent—this talk counts as an idealization (see Taylor, 2022a, as an account of how to move beyond this in tractable debate).

7 Alethic failures in cognitive science

What I have said in the preceding section leads to a general problem; namely, that in cognitive science there is no agreement about the scope of idealization even though there is agreement that idealization is taking place. If we aim to be "conscientious naturalists" (as I have argue above that we should), then this is highly problematic for the debate between strict-, non-, and quasi-factivism in this domain, because there is no agreement about descriptions of the relevant explanatory content in cognitive science. In short: when it comes to describing (the scope of) true or false explanatory content in cognitive scientific explanations, there is no agreement to be found (Doyle et al., 2019, 363). This is clearly disanalogous to the case of, say, the ideal gas law.

As a result, we are unable to identify which judgements about the idealizations of cognitive science are true, because there is no agreement about the idealizations in the first place. For example, we will be unable to determine if the judgement 'positing representations (of degrees of belief) allows us to make true claims about difference-making' is true. Thus, there is no way to appeal to cognitive scientific explanations to establish "extractive" strict- factivism or quasi-factivism, because there is no agreement about the scope of idealization in these explanations and, hence, no agreement about whether it is extracted true beliefs or our true judgements about idealizations that provide the authentic base of understanding (e.g. p^*) in this context.

And since we cannot agree on the scope of idealization in cognitive scientific explanations, there is also no way to establish non-factivism, because there is no way to establish the claim that false cognitive scientific explanations can provide understanding. This follows because we do not agree about which aspects of these explanations are false. For example, in the case of Bayesian cognitive science, we cannot agree on which part of the explanatory content is idealized (and so false) (cf. Mandelbaum, 2019; Mandelbaum et al., 2020). And without being able to agree in this regard, we cannot hope to establish the non-factivist claim that false explanations provide understanding.

Still, one may argue that progress is possible if strict-, non-, and quasi-factivists can agree among themselves about the scope of idealization in cognitive science. But this line of reasoning is confused, because those who come to agree on the scope of idealization in cognitive science cannot see themselves first and foremost as strict-, non-, or quasi-factivists, because then there is no guarantee that the descriptions of cognitive scientific explanations they favour are not biased towards their preferred position in the facticity debate.[11] What we require, then, is an impartial agreement on the scope of idealization in cognitive science.

The surest route to an impartial agreement on the scope of idealization in cognitive science is by undertaking a normative analysis of cognitive science. But this is exactly what philosophers of cognitive science are in the business of doing and, as I have shown above, the conclusions of this normative, analytic project thus far are clear: there is no agreement about the scope of idealization in cognitive science. The point, therefore, is that there is no robust, impartial consensus for the facticity debate to work from, and we cannot accept a consensus that strict-, non-, and quasi-factivists have establish for themselves.[12]

At present, this applies to descriptions of all explanations of cognition, because there is yet no agreement about which aspects of those explanations are idealized. To some, this argument may seem much too quick. But my recommendation to them is take a closer look at the relevant (philosophy of) science, where ongoing and fractious debates about, e.g., positing representations (cf. Chemero & Silberstein, 2008; Nanay, 2019), explanatory autonomy (cf. Churchland, 2013; Weiskopf, 2017), and the nature of cognition itself (Kaplan & Craver, 2011; Meyer, 2020) continue unabated. The point is simple: given the lessons of contemporary cognitive science, the agreement required to uphold conscientious naturalism is nowhere to be found.

8 Afactivism about understanding in cognitive science

I have argued that we cannot (yet) hope to establish either strict-, non-, or quasi-factivism by appeal to the relevant cognitive science, because there is (as yet) no agreement about the scope of idealization in cognitive scientific explanations. As a result, I claim that the conscientious naturalist should (for now) abandon alethic views of understanding in cognitive science and provisionally defend *afactivism*, which holds that:

Afactivism about Understanding Cognition Cognitive scientific explanation p provides understanding of x iff p is accepted, regardless of the truth or falsity of p.

[11] Note, however, that I am not claiming that strict-, non-, and quasi-factivists must necessarily disagree about the correct descriptions of cognitive scientific explanations, but only that this disagreement must be resolved prior to engaging in the dispute between strict-, non-, and quasi-factivism.

[12] As further support for this point, consider an ardent factivist, who could, *contra* any agreement between strict-, non-, and quasi-factivists, take the view that successful scientific explanations involve no idealization at all.

In the terms presented in Section 1, afactivism holds that, *in cognitive science*, in all cases where (1) is true:

$$\forall p \exists x (Ap \leftrightarrow U(x, p)) \tag{e}$$

Here, the use of "all" (e.g. $\forall$) is provisional, because it *might* eventually be possible to arrive at a consensus about the correct description of (some of) the relevant explanatory content in cognitive science. In this scenario, afactivism would be undermined, because it would be possible to refer to those descriptions to get to work establishing either strict-, non-, or quasi-factivism (as is putatively the case with the ideal gas law). This is exactly what conscientious naturalism would dictate, but, as I have shown, this agreement is nowhere to be found in cognitive science at present and so we will be unable to establish alethic views of understanding in that domain.

The key idea of afactivism is that we think of the understanding provided by cognitive scientific explanations as having a functional role that does not depend on those explanation's truth or falsity, but only on their being accepted. One open question is what it means to *accept* an explanation. Here, I follow van Fraassen (1980, 88) by taking that view that acceptance involves:

> a commitment to the further confrontation of new phenomena within the framework of that theory, a commitment to a research programme, and a wager that all relevant phenomena can be accounted for without giving up that theory.

Furthermore, I follow Doyle et al. (2019) in making a distinction between acceptance and belief. On their non-factivist view:

> the relevant difference is that it is sometimes rational to accept propositions known to be false if including these propositions among one's premises serves certain context-specific purposes, while it is never rational to believe propositions known to be false.

The afactivist will accept an almost identical claim, albeit with all talk of falsity (or truth) removed. So, for the afactivst, it will be rational to accept propositions with an unknown truth-value if including these propositions among one's premises serves certain context-specific purposes, while it will never rational to believe propositions with an unknown truth-value.[13]

The purposes I have in mind here may be pragmatic or practical. For instance, understanding a cognitive scientific explanation may help us to use that explanation to make better predictions, refine problems to be solved, make better specifications of the explananda, or control some target system.[14]

[13] Note that this does not imply that afactivists will never believe any explanation, because all scientific explanations will be constituted by sets of propositions of which some—e.g. human beings have cognitive systems—will have a determinate truth-value. It will be perfectly rational for the afactivist to believe in these aspects of any explanation.

[14] There is no space here to provide a detailed theory of purposive rationality as applied to acceptance. But there are some features that we can assume that any viable theory must possess. For example, that it is not rational to accept an inconsistent set of propositions for the same purpose. When trying to predict or control the behaviour of an ideal gas, for instance, it would not be rational to accept the proposition that particles in an ideal gas do not interact if one also accepts the proposition that ideal gases are constituted by interacting particles.

As an example, consider Bayesian cognitive science yet again. We have seen already that there is no agreement about the extent of idealization in Bayesian exaplanations of cognition and, hence, that there is no agreement about what is true or false about these explanations. However, Bayesian explanations have still been said to:

> provide a unifying framework for explaining the inferences that people make in different settings (Griffiths et al., 2010, 36).

And it has been argued that:

> the primary 'added value' of these [kinds of Bayesian] models is that they bring perception, action, and attention into a single unifying framework. They thus constitute the perfect explanatory partner [...] for recent approaches that stress the embodied, environmentally embedded, dimensions of mind and reason (Clark, 2013a, 201).

Clearly, understanding is involved here; for instance, understanding of unified phenomena or explanatory relationships. But what matters for this understanding is not the extent of idealized explanatory content, but whether the explanations are accepted and useful. That is, whether they have "value."

In the context of cognitive science, therefore, the advantage of afactivism is clear: it allows us to do what alethic views—i.e. strict, non-, and quasi-factivism—cannot; namely, take a stand on the understanding provided by cognitive scientific explanations even when we are faced with explanations that lack determinate truth-values. This is the case because the question of whether a cognitive scientific explanation provides understanding will depend only upon whether that explanation is accepted and used. And whether an explanation is accepted and used is, ultimately, an empirical question to be decided by studying the practices of cognitive science itself.

Afactivism, then, is not curtailed by conscientious naturalism in the same way as strict-, non-, and quasi-factivism, because the need to describe explanatory content is downplayed. Instead, the focus is on cognitive scientific practice itself; that is, the practice of accepting and using cognitive scientific explanations. Of course, the study of cognitive scientific practice may involve some normative analysis of its own, because such work may be necessary to establish when an explanation has been accepted and used, and in what way. But this analysis will be independent of the debate between strict, non-, and quasi-factivists, since the question of idealized content does not occur.

This appeal to scientific practice is the only science-based methodology that can gain traction *at this time* when considering understanding in cognitive science, because the science itself is rife with uncertainty about the alethic status of its explanations. But this should not cause us to abandon talk of understanding nor to commit ourselves to an alethic view that cannot yet be naturalistically justified. Rather, we should endorse afactivism and tie our talk of scientific understanding to non-alethic criteria. A good naturalist is left with no other choice.

9 Afactivism, facts, and problems

Afactivism about understanding cognition is the view that a cognitive scientific explanation p provides understanding of x iff p is accepted, regardless of the truth or falsity of p. For afactivism, then, the focus is on acceptance, which is constrained by rationality and which should be in the service of context-sensitive purposes. Still, one may wonder how afactivism differs from other theories of understanding that have recognised the problems that accrue when we tie understanding to truth. For example, Elgin's (2007, 41) "non-factive account in which felicitous falsehoods figure in understanding by exemplifying features that they share with facts."

On closer inspection, however, afactivism differs from Elgin's account in some crucial respects. Elgin (2007 33, 35) states that "Pretty plainly, understanding somehow answers to the facts" and that "The issue that divides factivists and non-factivists is not whether understanding must answer to the facts, but how it must do so." And this focus on facts is important, because it illustrates that Elgin's account is still operating within an alethic paradigm. This is most perspicuous in Elgin's (2004, 116, my italics) discussion of acceptance, where she argues that "To accept that p [$\cdots$] is to take it that p's divergence from truth, if any, is negligible. In that case, p is *true enough*" (Elgin, 2004, 116) (see also Elgin, 2017).

Thus, for Elgin (2004, 116), "understanding is not indifferent to truth," because the justification for idealizations in science is that they "figure in accounts that make sense of the facts." On Elgin's (2007, 38) account, therefore, understanding is "grounded in fact," since idealizations are taken to afford understanding only insofar as they furnish "epistemic access to matters of fact that are otherwise difficult or impossible to discern." Afactivism makes no such restriction about the capacity of idealizations to provide understanding, because afactivism eschews all talk of facts (and truth/falsity), since such talk results in an impasse when we cannot agree on the correct descriptions of idealized content.

Now, Elgin may worry that this "make[s] the world safe for postmodernist claptrap" by permitting understanding from "flagrantly false" explanations (I will return to this concern in the section on objections below). But this worry only makes sense if one insists on viewing the issue through the lens of alethic theories of understanding (hence the worry about *false* explanations providing understanding). Afactivism is the explicit—if provisional—rejection of this perspective (in cognitive science), which follows from the recognition that all present talk about the (scope of the) true/false content of cognitive scientific explanations is contested.

Still, one may insist that understanding ought to be tethered to the facts (and, indirectly, to truth). But, as I have explained above, in some domains (e.g. cognitive science) this claim—which amounts to a stubborn defence of the alethic paradigm—leads to an impasse and to the rejection of conscientious naturalism. Moreover, there are open questions about what facts *are* and even if we should permit facts into our ontology at all (see Betti, 2015). And, insofar as talk of facts is tied to truth, there are further questions about what truth *is*, as exemplified by long-standing and highly contentious debates between, say, correspondence (David, 2018), coherentist (Young, 2001), and contextual (Chang, 2012; Massimi, 2018) theories of truth.

It is notable that anti-alethic positions have already been developed in other domains. Consider, for example, discussions of scientific success and/or progress (of which achieving understanding may be a part). In that context, Kuhn (1970, 164) argues that scientific progress has nothing to do with truth, but is rather the increase of "both the effectiveness and the efficiency with the group as a whole solves new problems." Similarly, Laudan (1981, 145) argues that "science progresses just in case successive theories solve more problems than their predecessors" and Shan (2022, 745) argues that "Science progresses if more useful research problems and their corresponding solutions are proposed."[15]

These anti-alethic, "functional" theories of scientific progress are based on the acceptance of problem-definitions and problem-solutions. Afactivism about understanding is also based on acceptance that is rational and serves context-sensitive purposes. It is clear that problem-defining and problem-solving may be one such purpose, but afactivism does not restrict purposes to problem-defining and problem-solving alone. Other purposes may include, for example, the making of predictions or controlling relevant systems (see Taylor, 2022b for a discussion of how theories of concepts might support such purposes). As such, despite clear affinities, afactivism about understanding is a more permissive anti-alethic position than the anti-alethic positions developed to account for scientific progress.

10 Objections and replies

There are three obvious concerns with afactivism. The first is that afactivism pulls apart understanding and truth, and so leaves open the possibility that understanding is a mere subjective experience. In response, it is worth stressing that afactivism is a view about *understanding* and not about *truth*. In fact, afactivism remains neutral on the correct theory of truth. As such, the afactivist need not deny that there are absolute truths (although they could), because an afactivist could allow for two contradictory explanations to both provide understanding, while still maintaining that only one of the two explanations is (approximately) true.

Moreover, the concern that afactivism takes understanding to be mere subjective experience unfounded. In fact, the afactivist can agree with de Regt (2009a, 587) (and, e.g., Trout (2007)) that:

> the psychological sense of understanding has no epistemic function: a subjective feeling of understanding is neither necessary nor sufficient for scientific understanding of a phenomenon.

[15] There are different accounts of what problem-solving entails. According to Shan (2022, 746), "Kuhn suggests that puzzle-solving is an activity of looking for a solution which is sufficiently similar to a relevant paradigmatic problem-solution, while Laudan argues that a problem P is solved by a theory T if T entails an approximate statement of P." In contrast, Shan argues that we can, at best, specify a "common recipe" for problem-solving, which involves intertwined activities such as experimentation, conceptualisation, hypothesization, and reasoning.

For the afactivist, understanding has a *non-alethic* functional role that does not depend on the truth or falsity of explanations. However, this functional role is still tied to purposes and not mere subjective feeling, and may even be epistemic if we do not think that truth has exclusive value in our epistemic lives (cf. David, 2005).

The second obvious concern is that afactivism makes understanding relative to accepted explanations, such that one individual may gain understanding by accepting an explanation, but another individual may lack understanding by not accepting the same explanation. However, it is already widely accepted that the intelligibility and use of an explanation are crucial for achieving understanding. In this vein, de Regt (2009a, 595) claims that:

> of two people who possess exactly the same theories and background knowledge, one may achieve understanding of a phenomenon while the other does not. Accordingly, scientific understanding is not completely objective, that is, it is not fully independent of the subject.

I follow this line of reasoning and hold that worries of relativism are unfounded, since:

> Skills are acquired within a community and assessed by that community, which guarantees that judgments based on skills are not arbitrary.

In the case of afactivism, the relevant "judgments based on skills" are, of course, judgements of acceptance. For example, without adequate skill in the mathematics of imaginary numbers, one is unlikely to be able to accept (and understand) that $i^2 = -1$.

The final concern is that afactivism does not directly challenge strict-, non-, or quasi-factivism, because there must be a fact of the matter about whether understanding is provided by extracted true beliefs, false explanations, or true judgements about idealization, even if we tie understanding to acceptance. But this concern is, again, misplaced, because I have not argued for afactivism by claiming that strict-, non-, or quasi-factivism are incorrect. Rather, I have suggested that we must provisionally defend afactivism in the context of cognitive science until we reach a consensus about the scope of idealization, because such agreement is required to establish either strict, non-, or quasi-factivism.

Thus, my claim is that afactivism should be endorsed for the time being by those who prioritise the lessons of contemporary cognitive science over speculations about what lessons a mature cognitive science will disclose. Of course, strict-, non-, and quasi-factivists may argue that they are not worried by our inability to conclusively identify idealizations at this time, but to take this view is to turn one's back on conscientious naturalism as a philosophical methodology. So for those who genuinely aim to to be conscientious naturalists about our understanding of cognition, afactivism is the only option available.

11 Understanding cognition and understanding elsewhere

My argument in this paper has been that, for now, we should abandon alethic views of understanding and provisionally defend afactivism *in cognitive science*. However, I

have not argued that we should provisionally defend afactivism elsewhere. The reason for this is straightforward: our provisional defence of afactivism only makes sense when we fail to agree on the descriptions of explanatory content and alethic views of understanding cannot gain traction, but such disagreement is not apparent in all—or even most—domains.[16] In the context of thermodynamics, for instance, all agree on the idealized content of the ideal gas law.

I am not, therefore, putting forward afactivism as a general theory of scientific understanding and I accept that even while one provisionally defends afactivism in cognitive science, one may also defend another view (e.g. strict-, non-, or quasi-factivism) in another domain (e.g. thermodynamics). It is an empirical question whether or not there is agreement about the content of explanations in any given domain, which is to be addressed by, among other things, analysing the explanations produced in that domain and attending to the practices undertaken in the domain in question.

Now, a strict-, non-, or quasi-factivist might argue that their theories of understanding are general theories of scientific understanding, and so when we establish strict-, non-, or quasi-factivism in one domain, we establish that theory for all domains. Thus, strict-, non-, or quasi-factivists might argue that if we can establish either view by referring to, say, the ideal gas law, then we establish that view *tout court*. This argument, however, rests on assumption about the unity of science (and reality) that many do not share.

Consider, for example, the view that science is disunifed, because higher-level theories cannot always be derived from lower-level theories (Fodor, 1974). Furthermore, consider the view that reality itself is disunified, because the world is somehow "dappled" (Cartwright, 1999) or "disordered" (Dupre, 1993). This kind of ontological pluralism often holds that "there is not only one correct way of carving up the natural world and that different scientific interests and values lead to different but equally valid ontologies" (Ludwig & Ruphy, 2021), and has been defended as the proper epistemic attitude to adopt, because of the complexity of the world or because of the world's lack of an ordered structure (Longino, 2006, 2013; Waters, 2017).

While strict-, non-, and quasi-factivists may be quick to reject such views, this would be unsurprising, because it would allow them to side-step the issues I have raised in this paper about problematic domains—e.g. cognitive science—where no agreement about the descriptions of explanatory content can (yet) be found. For the afactivist, however, there is no good reason to take a stand on the (dis)unity of science or reality at this time, because there is no pressing need to uphold a view on the general applicability of strict-, non-, or quasi-factivism.

Thus, the afactivist can adhere to "metaphysical abstinence" in the vein of Ruphy (2005, 118), who argues that:

> philosophical questionings of local scientific methodologies, when based on empirical considerations, are undoubtedly valuable for scientists, but they are much less valuable when based on metaphysical contentions. [...] If philosophers want their arguments to remain relevant for scientific practice, I am afraid that

[16] See Taylor and Williamson (2022) for an account of another kind of disagreement in cognitive science: disagreement about evidence.

they must be more modest about the status of these arguments: only temporally qualified arguments that stay clear of metaphysical contentions can usefully bear on discussions of the merits and limits of reductionist approaches in science.

The idea, then, is that by taking up afactivism we can remain neutral about the (dis)unity of science and reality, just as we can remain neutral about whether strict-, non-, or quasi- factivism is correct in those domains in which we fail to agree on the descriptions of the relevant explanatory content. This neutrality has value, because it ensures that the debate about scientific understanding does not get ahead of the science itself. But the picture that falls out of this neutral stance is, for the time being at least, one of a patchwork of different routes to understanding; some instances, perhaps, fitting a strict-, non-, or quasi- factivist model, but others better described in afactivist terms.

Funding Funded by The Leverhulme Trust.

References

Baylor, D. A. (1987). Photoreceptor signals and vision. Proctor lecture. *Investigative Ophthalmology & Visual Science, 28*(1), 34–49.

Betti, A. (2015). *Against facts*. MIT Press.

Cartwright, N. (1999). The dappled world: A study of the boundaries of science. Cambridge University Press

Chang, H. (2012). Is water h2o?: Evidence, realism and pluralism (Vol. 293). Springer

Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences, 34*(4), 194–196.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. USA: Oxford University Press.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of science, 75*(1), 1–27.

Churchland, P. M. (2013). *Matter and consciousness*. Cambridge, MA: MIT press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Be-havioral and Brain Sciences, 36*(3), 181–204.

Coelho Mollo, D. (2020). Deflationary realism: Representation and idealisation in cognitive science. Mind & Language.

Colombo, M., & Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science, 68*(2), 451–484.

David, M. (2005). Truth as the primary epistemic goal: A working hypothesis. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 363–377). Wiley-Blackwell.

David, M. (2018). The correspondence theory of truth. In The oxford handbook of truth (pp. 219–237, Vol. 1). Oxford University Press.

De Regt, H. (2009). The epistemic value of understanding. *Philosophy of Science, 76*(5), 585–597.

de Regt, H. (2009). Understanding and scientific explanation. In Scientific understanding: Philosophical per-spectives (pp. 21-42). University of Pittsburgh Press.

Doyle, Y., Egan, S., Graham, N., & Khalifa, K. (2019). Non-factive understanding: A statement and defense. *Journal for General Philosophy of Science, 50*(3), 345–365.
Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.
Egan, F. (2014). How to think about mental content. *Philosophical Studies, 170*(1), 115–135.
Elgin, C. (2004). True enough. Philosophical. *Issues, 14*(1), 113–131.
Elgin, C. (2007). Understanding and the facts. *Philosophical Studies, 132*(1), 33–42.
Elgin, C. (2009). Is understanding factive. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic value* (pp. 322–330). Oxford University Press.
Elgin, C. Z. (2017). *True enough*. MIT Press.
Fodor, J. A. (1974). Special sciences. *Synthese, 28*, 97–115.
Fusco, G., Ciccarone, S., Petrucci,M., Cozzani, B., Vercelli, G., Cotugno, A., & Bufalari, I. (2023).Altered processing of conflicting body representations in women with restrictive anorexia nervosa. *Psychological Research*, 1-14.
Greco, J. (2013). Episteme: Knowledge and understanding. In K. Timpe & C. A. Boy (Eds.), *Virtues and their vices* (pp. 285–301). Oxford University Press.
Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilisticmodels of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences, 14*(8), 357–364.
Grimm, S. (2006). Is understanding a species of knowledge? *The British Journal for the Philosophy of Science, 57*(3), 515–535.
Grimm, S. (2012). The value of understanding. Philosophy. *Compass, 7*(2), 103–117.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175.
Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge, MA: MIT Press.
Ishikawa,M., & Senju, A. (2023). Action value calculations in social context from infancy. *Trends in Cognitive Sciences*.
Kaplan, D., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of science, 78*(4), 601–627.
Korenbrot, J. I. (2012). Speed, sensitivity, and stability of the light response in rod and cone photoreceptors: Facts and models. *Progress in Retinal and Eye Research, 31*(5), 442–466.
Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
Kvanvig, J. L. (2009). Précis of the value of knowledge and the pursuit of understanding. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic value* (pp. 309–313). Oxford University Press.
Laudan, L. (1981). A problem-solving approach to scientific progress. In Scientific revolutions (pp. 144-155). Oxford University Press.
Lawler, I. (2021). Scientific understanding and felicitous legitimate falsehoods. *Synthese, 198*(7), 6859–6887.
Li, A. Y., Yuan, J. Y., Pun, C., & Barense, M. D. (2023). The effect of memory load on object reconstruction: Insights from an online mouse-tracking task. *Attention, Perception, & Psychophysics*, 1-19.
Li, P. H., Field, G. D., Greschner, M., Ahn, D., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., & Chichilnisky, E. J. (2014). Retinal representation of the elementary visual signal. *Neuron*, *81*(1), 130–139.
Longino, H. (2006). Theoretical pluralism and the scientific study of behavior. In Scientific pluralism (pp. 102-131). University of Minnesota Press.
Longino, H. (2013). Studying human behavior: How scientists investigate aggression and sexuality. University of Chicago Press.
Ludwig, D., & Ruphy, S. (2021). Scientific Pluralism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Winter 2021.* Metaphysics Research Lab: Stanford University.
Machery, E. (2009). *Doing without concepts*. New York: Oxford University Press.
Mandelbaum, E., Won, I., Gross, S., & Firestone, C. (2020). Can resources save rationality? 'Anti-Bayesian' updating in cognition and perception. *Behavioral and Brain Sciences*, 143.
Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language, 34*(2), 141–157.
Massimi, M. (2018). Four kinds of perspectival truth. *Philosophy and Phenomenological Research, 96*(2), 342–359.

Meyer, R. (2020). The non-mechanistic option: Defending dynamical explanations. *The British Journal for the Philosophy of Science*.
Mizrahi, M. (2012). Idealizations and scientific understanding. *Philosophical Studies, 160*(2), 237–252.
Nanay, B. (2019). Entity realism about mental representations. *Erkenntnis*, 1-17.
Odenbaugh, J. (2011). True lies: Realism, robustness, and models. *Philosophy of Science, 78*(5), 1177–1188.
Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.
Prinz, J. J. (2002). Furnishing the mind: Concepts and their perceptual basis. MIT Press.
Rancourt, B. T. (2017). Better understanding through falsehood. *Pacific Philosophical Quarterly, 98*(3), 382–405.
Rice, C. (2016). Factive scientific understanding without accurate representation. *Biology & Philosophy, 31*(1), 81–102.
Roska, B., & Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature, 410*(6828), 583–587.
Ross, L. (2023). The truth about better understanding? *Erkenntnis, 88*(2), 747–770.
Ruphy, S. (2005). Why metaphysical abstinence should prevail in the debate on reductionism. *International Studies in the Philosophy of Science, 19*(2), 105–121.
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
Shan, Y. (2022). A new functional approach to scientific progress. *Philosophy of Science, 86*(4), 739–758.
Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.
Soo, F. S., Schwartz, G. W., Sadeghi, K., & Berry, M. J. (2011). Fine spatial information represented in a population of retinal ganglion cells. *Journal of Neuroscience, 31*(6), 2145–2155.
Strevens, M. (2011). Depth: An account of scientific explanation. Harvard University Press.
Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A, 44*(3), 510–515.
Strevens, M. (2017). How idealizations provide understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New essays in epistemology and the philosophy of science* (pp. 37–49). Routledge.
Taylor, S. D. (2022). *Cognitive instrumentalism about mental representations. Pacific PhilosophicalQuarterly, 103*(3), 518–550.
Taylor, S. D. (2022). Concepts as a working hypothesis. *Philosophical Psychology, 35*(4), 569–594.
Taylor, S. D., & Sutton, P. R. (2021). A frame-theoretic model of bayesian category learning. In S. Löbner, T. Gamerschlag, T. Kalenscher, M. Schrenk, & H. Zeevat (Eds.), *Concepts, frames and cascades in semantics, cognition and ontology* (pp. 329–349). Springer International Publishing Cham.
Taylor, S. D., & Vosgerau, G. (2021). The explanatory role of concepts. *Erkenntnis, 86*(5), 1045–1070.
Taylor, S. D., & Williamson, J. (2022). Evidence and cognition. *Erkenntnis*. https://doi.org/10.1007/s10670-022-00611-0
Thomson, E., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines, 28*(1), 191–235.
Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69*(2), 212–233.
Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass, 2*(3), 564–591.
Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
Wandell, B. A. (1995). *Foundations of vision*. Sinauer Associates.
Waters, C. K. (2017). No general structure. In Metaphysics and the philosophy of science: New essays (pp. 81-108). Oxford University Press.
Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy, 104*(12), 639–659.
Weiskopf, D. A. (2017). The explanatory autonomy of cognitive models. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 44–69). New York, NY: Oxford University Press.
Young, J. O. (2001). A defence of the coherence theory of truth. *Journal of philosophical research, 26*, 89–101.
Yu, S., Stock, A., Münchau, A., Frings, C., & Beste, C. (2023). Neurophysiological principles of inhibitory control processes during cognitive flexibility. *Cerebral Cortex*, bhac532.
Zach, R. (1978). Selection and dropping of whelks by northwestern crows. *Behaviour, 67*(1–2), 134–147.