Synthese (2024) 203:48
https://doi.org/10.1007/s11229-023-04445-3

ORIGINAL RESEARCH

Don't get it wrong! On understanding and its negative phenomena

Haomiao Yu[1] · Stefan Petkov[2]

Received: 5 June 2023 / Accepted: 3 December 2023

Abstract
This paper studies the epistemic failures to reach understanding in relation to scientific explanations. We make a distinction between genuine understanding and its negative phenomena—lack of understanding and misunderstanding. We define explanatory understanding as inclusive as possible, as the epistemic success that depends on abilities, skills, and correct explanations. This success, we add, is often supplemented by specific positive phenomenology which plays a part in forming epistemic inclinations—tendencies to receive an insight from familiar types of explanations. We define lack of understanding as the epistemic failure that results from a lack of an explanation or from an incorrect one. This can occur due to insufficient abilities and skills, or to fallacious explanatory information. Finally, we characterize misunderstanding by cases where one's epistemic inclinations do not align with an otherwise correct explanation. We suggest that it leads to potential debates about the explanatory power of different explanatory strategies. We further illustrate this idea with a short meta-philosophical study on the current debates about distinctively mathematical explanations.

Keywords Explanatory understanding · Epistemic inclination · Explanatory pluralism · Misunderstanding · Disputes in science and philosophy · Distinctively mathematical explanations

✉ Stefan Petkov
yaggdrasil@yahoo.com

Haomiao Yu
yuhaomiao812@gmail.com

[1] Department of Philosophy, University of Guelph, Guelph, ON N1G 2W1, Canada

[2] School of Philosophy, Beijing Normal University, Room A801 Front Main Building, No. 19 Xinjiekouwai Str, Haidian District, Beijing, China

Published online: 30 January 2024

1 Introduction

Recent analysis of understanding has shown that it comes in degrees (Baumberger, 2019; Khalifa, 2017). As such, understanding and its negatives form a contrary, not a contradictory category. Thus, mapping the spectrum of understanding is not a trivial task. For instance, some accounts have offered suggestions about the evaluation of degrees of understanding (Baumberger, 2019; Kelp, 2015; Khalifa, 2017; Petkov, 2021), and specified some of the conditions of understanding, approximate understanding, and the failures to achieve understanding (Park, 2017, 2020). Given this rich literature and its recent developments, we believe that we can present an updated and unified account focused on explanatory understanding, and then use such an account to present a more systematic and expanded study of the conditions under which explanatory understanding cannot arise. The conditions of such failures to achieve understanding are of primary interest to us, and their analysis is the central goal of the present study.

Here we limit our analysis to cases of understanding that depend on explanations, because both the logic and pragmatics of explanations as well as the general features of explanatory understanding are all well mapped (Khalifa, 2017). Therefore, for the most part we will base our analysis on cases that can be covered by classical accounts of explanations. This narrow scope might initially seem limited, but we believe that having a clear foundational analysis of explanatory understanding and the conditions under which it cannot arise, will offer a solid starting point for discussing more complex scenarios, such as cases of explanatory understanding derived from models that assert a level of idealization. Therefore, we focus on explanations based on classical theories of explanations in the main text. In the appendix, we shall suggest how our analysis can be extended to broad considerations about understanding involving modeling and idealization which have been hotly debated in recent literature.

To warrant the claim that an analysis of the failures to reach understanding is philosophically and methodologically interesting, we distinguish between genuine understanding,[1] and two types of epistemic failures: lack of understanding and misunderstanding. Whilst some of the existing literature has touched on lack of understanding (Khalifa, 2013, 2017; Park, 2017), no analysis of the difference between misunderstanding (as we define it) and the lack of understanding yet exists.

In Sect. 2, we roughly chart explanatory understanding, based on existing accounts, as requiring subjective and objective components. We describe its objective component in terms of a correct explanation in Sect. 2.2. This correctness is often determined by specific theories of explanations that cover structural, informational, and ontological requirements for explanations. The subjective component for us can be captured by the cognitive abilities and skills necessary to grasp or construct a correct explanation, plus the concomitant phenomenon of insight (the "aha!" moment) or a gradual increase of satisfaction (Sect. 2.3). Finally, we frame explanatory understanding as the alignment between the subjective and objective components (Sect. 2.4).

[1] In what follows, we use the terms "understanding" and "genuine understanding" interchangeably to designate the same epistemic success.

Using such a general theory of explanatory understanding, in Sect. 3 we systematize the existing claims about the *lack of understanding*. These concern situations where the explanation is incorrect, or situations where the cognitive abilities and skills of the explanation seekers are insufficient to grasp or construct a correct explanation.

We suggest an addition to the existing views on explanatory understanding. We hypothesize that the epistemic success of understanding gives rise to epistemic inclinations. We describe an epistemic inclination as the subjective tendency of the habitual employment of previously proven to be successful, explanatory strategies. This addition serves to highlight our characterization of misunderstanding.

In Sect. 4, we offer (what we believe) is our original contribution to the literature: An analysis of the notion of misunderstanding. This concerns cases where one's epistemic inclinations do not align with an otherwise correct explanation. Therefore, misunderstanding is distinguished from a lack of understanding by the rejection of a correct explanation in the presence of skills and abilities necessary for its grasping.

In Sect. 5, we illustrate the cases of misunderstanding by offering a metaphilosophical analysis of the debate around the existence of distinctively mathematical explanations.

In Sect. 6, we conclude with the idea that the presence of epistemic inclinations in scientific or philosophical debates does not necessarily lead to dogmatism.

Finally in the Appendix, we offer some suggestions on how our analysis can be extended to more complex cases, such as explanations that contain idealizations, as well as understanding that seems to directly stem from modeling without the presence of an explanation.

2 Explanatory understanding

2.1 The subjective and objective components of explanatory understanding

Recent work on understanding shows that it is a broad epistemic category (Baumberger et al., 2017). Thus, an exhaustive analysis of understanding and its negatives would extend beyond the scope of a single paper, or at the cost of analytic precision. To avoid this, here we aim to obtain a good sketch of explanatory understanding and the conditions under which it cannot occur. This will enable us to use it as a starting point for developing a more comprehensive picture in the future.

Another difficulty is that, besides the numerous discussions of different sources of understanding, the concept of understanding itself is hotly debated too. In order to avoid sidetracking, we strive to provide an account that is maximally compatible with the existing views. To achieve this, our first step is to split the notion of understanding as having two components. Typically, explanatory understanding is seen as requiring the right kind of explanations and the right kind of epistemic subjects (Kelp, 2015).

Therefore, its components can be presented as objective and subjective ones.[2] Using this distinction, our first sketch at a definition of explanatory understanding is:

> The epistemic achievement that results from (subjectively) grasping or constructing the (objectively) correct explanation for a given fact.

To clarify, (for more details see Sect. 2.3), subjective grasping involves the abilities and skills to "assimilate" the explanatory information or construct an explanation. How this assimilation can be assessed, as for instance by the sort of inferential "tests" that Grimm (2010) or Newman (2014) posits, is something that we leave aside. Subjectively constructing an explanation for us simply means that the epistemic agent suggests an explanation to herself, or to another explanation seeker, or to a relevant scientific community.

What we call an objectively correct explanation, does not necessarily mean anything like the realist claim that the world contains explanations for its facts, and these need only to be verbalized. By "objectively correct" here we merely mean that an explanation is a type of discursive artifact—an artifact that serves a communicative purpose (Potochnik, 2017; Wright, 2012).

As such, once the explanatory seeking problem is formulated clearly (i.e., conceptualized as a specific why-question), and the explanation is presented, we are left with a linguistic or conceptual construct of some kind. In the simplest case, this structure involves a question and an answer (van Fraassen, 1980); but more often in scientific contexts, it is represented by some sort of inference (Kitcher, 1989; Petkov, 2015). Consequently, this discursive construct can be evaluated in terms of its structural and informational properties.[3] Such an evaluation can be made from the first-person, reflective perspective of the creator of the explanation, or from another—the explanation seeker. Even more broadly, an evaluation of an explanation might come from a whole epistemic group (i.e., a scientific community). All in all, the epistemic success of understanding can be seen as an alignment between these subjective and objective components.[4]

To further clarify on these, we first focus on what correct explanations are in Sect. 2.2, and then move on to what grasping or constructing such explanations entails in Sect. 2.3.

2.2 Explanations and the objective components of understanding

In this section, we analyze the requirements for explanations from two sides: theories of explanations, and theories of understanding. Starting with the theories of explanations, our goal here is not to develop an original theory of explanation. Therefore we assume explanatory pluralism as a foundational claim. On this view, there is no dominant way

[2] We believe that this categorization is warranted, because the contemporary analysis of explanatory understanding has proven, contrary to the received view (Hempel, 1965; Kitcher, 1989; Woodward, 2003), that only analyzing the structural properties of explanations is not sufficient to arrive at a notion of understanding. The involvement of subjects and their cognitive abilities is inevitable (Khalifa et al., 2023).

[3] Such as the explanatory relation between the explanans and the explanandum, the relevance of the explanatory information and its evaluation.

[4] For a clarification on the notion of alignment see Sect. 2.4.

in which scientific explanations are successful in producing understanding. Moreover, each theory of explanations has its own requirements (Khalifa et al., 2020; Pincock, 2018; Rice & Rohwer, 2021), for what counts as a correct explanation. In addition to this plurality of explanatory approaches, recent theories of understanding argue that there are different criteria for evaluating the content of explanations (de Regt, 2017, 2023; Khalifa, 2017, 2023).

Taking a hint from Khalifa (2017), we classify these requirements as:

1. *Structural requirements*, postulated by theories of explanations;
2. *Informational requirements*, postulated by theories of explanations;
3. *Ontological requirements* for the content of explanations, postulated by theories of understanding.

We characterize explanations as *correct*, when they meet all the criteria above; and *incorrect* when they fail to meet some. However, criterion 2 has a complication. It can be under- or over-satisfied. Explanations that under-satisfy criterion 2 count as *incomplete*, as they do not provide sufficient information for solving the explanatory seeking problem. Consequently, such explanations also count as *incorrect*, but this must be taken narrowly in relation only to the specific informational requirements of the explanatory seeking problem. On the other hand, explanations that over-satisfy criterion 2 count as *excessive*, as they provide more than sufficient information. By the same token, excessive explanations are *incorrect* as well.[5] The rest of this section is dedicated to clarifying these requirements.

Typically, theories of explanations provide two types of requirements for explanations. The first can be labeled as *structural.* It describes the form that the explanation should take, relevant to the type of explanatory relation it instantiates. The second can be labeled as *informational.* It encompasses the requirements for the type and the amount of information that the explanation should exhibit.

Regrettably, here we can only clarify these criteria through some rough, typical examples. We found it hard to provide a unified and completely explicit account, because each theory of explanations provides specific requirements for both the structural and informational properties of its type of explanations. To put it differently, explanatory pluralism postulates that different possibilities of explanations exist, because each theory of explanations suggests (often subtle) differences in both the structure and informativeness of explanations.[6]

Nevertheless, starting with *structural requirements*, these can be very minimal in scale, such as those discussed by van Fraassen (1980) who characterizes explanations as direct answers to why-questions. Typically, however, explanations take the form of an inference from the explanans to the explanandum. In such cases, the structural requirements overlap with the formal requirements for such inferences plus extra requirements for the type of explanations the theory covers. For instance, in deductive-nomological explanations (Hempel, 1965), explanations take the form of a deductive

[5] Importantly we suggest that their incorrectness and thus the failure or success of understanding depend on both the ability of the explanation seeker to reduce them to correct ones, and to their complexity which might hamper such a reduction.

[6] Rice and Rohwer (2021) have suggested that so long as an explanation covers an often-assorted cluster of properties associated with the stock of accepted theories of scientific explanations, that explanation can provide understanding.

inference, with a general premise that is a lawful generalization, plus relevant auxiliary conditions; and the explanandum must figure as a conclusion from them. In distinctively mathematical explanations (Lange, 2013), the structure is similar; except that the generalization is a description or a proof that some property is instantiated in a given mathematical structure. That structure then is linked with /or represented by/ some of the properties of the explanandum. Finally, the explanatory conclusion is that the property in the explaining mathematical structure holds also for the explanandum (See Sect. 5).

Causal explanations come in various types as well. They can take the form of a causal narrative (Currie & Sterelny, 2017), or a description of a mechanism (Craver, 2007; Craver & Darden, 2013), or a conditional (Lewis, 1973). In the case of counterfactual causal explanations for instance, the antecedent of the conditional is the cause (explanans), the consequent is the effect (explanandum), and the relation between the two is that of causal dependence (conceived as necessity or probability) (Lewis, 1973; Woodward, 2003).

Of course, explanations can also display complex structural properties by being both mathematical and causal (Craver & Povich, 2017). Some have suggested that explanations by unification can also be seen as cases of relating deductive and causal information (Petkov, 2015).

Despite this plurality, theories of explanations share the common ground that, if the structural requirements for an explanation are violated, the resulting explanation would be incorrect. For example, an incorrect deductive explanation could be an invalid deduction; or an incorrect causal explanation could postulate spurious causal links, etc.

These structural requirements are complemented by *informational requirements*, as theories of explanations also provide criteria for the informativeness of the explanans. This covers the right kind and amount of information in the explanation, relevant to the requirements of the specific explanatory seeking problem. Judging the right kind and amount of information depends on both the explanatory seeking problem and the type of explanation sought. When the informational requirement is violated, the resulting explanation can be *incomplete, excessive,* or simply *inadequate.* To paint a clearer picture, we provide an illustration inspired by van Fraassen's pragmatics (1980) and Potochnik's (2017) account of causal explanation patterns. Given a specific explanandum, the following three explanations can all count as *incomplete, excessive, sufficiently informative* or even *inadequate:*

(A) All salt dissolves in water, because CH: salt is sodium chloride, and sodium has loosely bound outer electrons and is therefore highly reactive;
(B) All salt dissolves in water, because CH, and STR: but a mass of granular soluble substances dissolves more quickly than a mass of cubes of the same substance.
(C) I bought salt cubes, because they were on sale.

Given an explanatory seeking problem "Why does *salt* dissolve in water?" *A* would be sufficiently informative, *B* excessively informative, and C inadequate. On the other hand, if the question is "Why does a cube of *this substance* dissolve more slowly than a heap with the same mass?", *A* would be incomplete, *B* sufficiently informative, and

C inadequate. Finally relative to "Why *do we have this salt that* dissolves so slowly?", *A* would be inadequate, *B* incomplete, and *C* sufficiently informative.

This framing of the objective components of explanations allows us to investigate explanations as clearly defined explanatory seeking problems, and adequately evaluate the informativeness of the explanatory answers. Consequently, this permits us to say that, at least in some situations once a clear explanatory seeking problem is proposed there are cases of informatively correct, informatively incomplete, or excessive explanations, relative to that problem.

Here we add that the structural and informational components of explanations are also intimately linked. This is because, very often scientific problems are open-ended. In this sense, it is possible to use more than one explanatory strategy to solve the same problem. A classic example is Salmon's balloon case (Salmon, 1990). Where the explanatory problem "Why did the balloon fly towards the cockpit of a plane during liftoff?", can receive a causal or a general theoretical (i.e., via unification) explanatory solution. However, in order for any of these solutions to be counted as correct, they essentially impose a structural constraint on the explanandum, by interpreting some of its aspects. This postulates a more concrete explanatory seeking problem. Consequently, specific causal or unificatory information would count towards a correct solution. Importantly each of these solutions is judged via their structural and informational properties, relevant to the theory of explanations, which covers them. We return to this issue in our discussion on understanding and inclinations (in Sect. 2.4). But for now, it is important to establish that the structural and informational requirements for an explanation taken together can lead to asserting a narrowly correct explanation, even where a plurality of other explanations can be suggested.

The often open-endedness of scientific problems also suggests that, the explanatory seeking problem can require a structurally complex explanation, that synthesizes structural and informational requirements from more than one theory of explanation. Such explanatory projects can thus impose their own constraints on what counts as structurally and informatively correct explanation. This can be illustrated again by the debate between the accounts of distinctively mathematical and causal explanations in our case study. As we shall see in Sect. 5, proponents of distinctively mathematical explanations claim that, for cases of distinctively mathematical explanations, the addition of causal information leads to excessively informative explanations; whilst proponents of causal accounts claim that distinctively mathematical explanations are incomplete, because a correct explanation should be a combination of mathematical and causal information.

To summarize, relative to the informational requirements of explanations, the sufficiency is determined in relevance to the specific explanatory seeking problem, and the theories of explanations that cover it. An explanation can be inadequately informative and thus incorrect; incomplete when the information is not sufficient; and excessive, when the information is more than sufficient.[7] Finally, an explanation can be correct when the structural criteria are met and the information in the explanation is sufficient.

Turning our focus to the *ontological requirements*, theories of understanding also dispute the criteria for judging the truthfulness of explanatory information. This has

[7] More than sufficient as in more than what is required for solving the specific problem.

been at the center of the so-called factivity debate. According to factivists, the ontological status of the information in an explanation should be true in a correspondence sense. This means that the truth makers of the information in the explanans must be facts (Wilkenfeld, 2017). Non-factive accounts (de Regt, 2017; de Regt & Gijsbers, 2017), on the other hand, claim that explanations can fall short of being true in this correspondence sense. As an alternative, they claim that the explanans can be true, based on a truth maker that is something other than a straightforward correspondence with facts.

To provide more details, Wilkenfeld (2017) holds that understanding is tied to truth in terms of representational accuracy, assuming a correspondence theory of truth. For non-factive accounts such as de Regt's (2015, 2017, 2023) pragmatic approach, the evaluation of the explanans can involve a variety of criteria, such as its intelligibility, effectiveness in promoting predictions, practical applications, or its general heuristic value (de Regt & Gijsbers, 2017). As such de Regt (2015), for instance, suggests that the evaluation of the explanans could still be true, while the truth maker is due to some other criterion, besides factual correspondence. For instance, he promotes the idea that a pragmatic theory of truth can function as such an evaluation (de Regt, 2015).

This debate has further spilled out to considerations about scientific realism vs instrumentalism, as well as the role of idealization in explanations (see the Appendix). To avoid sidetracking here we adopt the solution proposed by Khalifa (2023).

Given the plurality of judgments for evaluating the appropriate type of explanatory information, Khalifa (2023) has described the identification of truth makers as imposing further ontological requirements on explanations. In order to transcend the factivity debate, Khalifa has suggested a reconciliation between the two opposing positions, by promoting what he calls "explanatory voluntarism" (2023, p. 45). This means that theories of explanations provide the structural and informational criteria for explanations, whilst the ontological requirements on explanations coming from theories of understanding need not be committed. The ontological requirements can be a simple correspondence with facts for factivists, or a more complex set of criteria, such as the pragmatic ones promoted by non-factivists. According to Khalifa's explanatory voluntarism, both factivists and non-factivists can agree that correct explanations provide understanding, while disagreeing on the ontological requirements for correctness. Thus, for instance both camps would agree that explanations derived from idealized models provide understanding, but disagree on how to evaluate the ontological status of such information. For factivists these explanations are true or approximately true in the correspondence sense, whilst for non-factivists their evaluation depends on pragmatic truth or other contextual criteria (see the Appendix).

Without delving further into the factivity debate, we adopt Khalifa's explanatory voluntarism. This enables us to claim that when the ontological requirements for an explanation are not met, the explanation fails to be correct. In addition, we propose that both structural and ontological requirements are also critically involved in providing understanding via correct explanations. We believe that this move sets the bar for understanding from a theory-neutral perspective. It in turn advances our discussion of lack of understanding as resulting from incorrect explanations.

2.3 The subjective components of explanatory understanding

Here, we explicate the subjective dimension of understanding as the epistemic achievement (1) supplemented by a specific phenomenal hue, and (2) requiring cognitive abilities and skills. This setup has a theoretical motivation, as the literature mostly discusses the subjective side of understanding under the concept of grasping. Grasping in turn has two sides—phenomenological and epistemic.

Psychologically, grasping has been discussed as being accompanied by a specific phenomenal hue—the so-called "aha!" feeling or an insight moment (Baumberger et al., 2017). The literature converges on the idea that the subjective feeling of understanding is neither necessary nor sufficient for understanding (de Regt, 2009; Hills, 2009; Lipton, 2009; Trout, 2002; Ylikoski, 2009), due to the unreliable nature of the feeling itself. This is certainly the case for the typical analysis of understanding. But we do not think it shows that the phenomenological dimension of understanding is of no theoretical importance, especially when we consider explanations and the relevant epistemic outcomes such as understanding, lack of understanding, and misunderstanding. For instance (as we shall discuss in Sect. 4), the absence of the feeling of understanding can lead to rejecting an otherwise correct explanation, even when the cognitive abilities related to reaching understanding via such an explanation are present. This motivates us to include the phenomenological side of understanding into our analysis.

Understandably, most philosophical analyses of understanding in terms of grasping emphasize the epistemic side, whilst underestimating the phenomenological side. This is perhaps because the initial efforts were focused on clarifying what kind of epistemic success understanding is. But we believe that theories of understanding are now mature enough to be extended to provide a more detailed analysis of the phenomenological side of understanding. Also, it seems to us that at least some specific epistemic failures depend on such a phenomenology. Therefore, here we extract the claims on the epistemology of grasping and combine them with the phenomenology of understanding.

The epistemic side of grasping can be broadly characterized as an assimilation of the explanatory information into one's belief system. This assimilation has been described in numerous ways. For instance, Khalifa (2017) characterizes grasping as the agent's capacity to assimilate an explanation within her belief system which potentially contains other relevant explanations of the phenomenon. Others (Grimm, 2010; Newman, 2014) analyze grasping as the ability to use the explanatory information to construct inferences similar to the one exemplified by the explanation. Grasping can be embedded into theories of explanations as well. For instance, under the manipulationist accounts of causation, grasping is the ability to "see" how a manipulation of the explanans would lead to a change in the explanandum, etc. (Woodward, 2003).

All these accounts share the common ground that grasping requires a number of cognitive abilities and skills. However, they diverge on what such abilities and skills are, and if they provide a distinct epistemic achievement, different from knowledge. In other words, they try to answer the question: "Is grasping the correct explanation simply equivalent to knowing the correct explanation?".

This problem is at the center of the debate between knowledge-based accounts—which classify understanding as a type of knowledge (Grimm, 2006; Kelp, 2015; Khalifa, 2017; Mizrahi, 2012; Strevens, 2017), and skill-based accounts—that distinguish understanding as a distinct epistemic category (de Regt, 2017, 2023; Newman, 2014), characterized by the development of skills. In order to clarify the role of cognitive abilities and skills for grasping, we will shortly overview the debate between de Regt and Khalifa.

For the skill-based accounts championed by de Regt (2017, 2023), understanding is a distinct epistemic category, because it requires the development of context specific skills. These skills can be minimally the skills required to make the explanation intelligible to the explanation seeker. An explanation is intelligible to a subject if she can solve similar problems as the one presented by the explanation, or more broadly if she "sees" how the explanatory relation works. That is to say, for instance one's deductive reasoning skills are required to understand the explanandum by a deductive-nomological explanation, because they are at the heart of solving the explanatory seeking problem. For de Regt, such skills are separated from what is required for believing and thus simply knowing the correct explanation.

Opposing this view, the knowledge-based accounts downplay the importance of such skills. These accounts do not deny that some abilities and skills are involved in grasping. However Khalifa, for example, argues that these are of a rather trivial kind. This is because scientific knowledge already "requires a fair amount of ability" (Khalifa, 2017, p. 54). Moreover, according to him, these abilities are reducible to knowledge-that, and thus do not exceed knowing (Khalifa, 2012). Therefore, concerning the same case of deciding if a student understands the explanandum through a given explanation or is merely able to recite the explanation, Khalifa makes the same judgment. However, his assertion that the student lacks understanding is based on the fact that the student does not possess the knowledge of deduction, that concerns the inferential connection between the explanans and the explanandum (Khalifa, 2012, p. 26).

To clarify, for both de Regt and Khalifa, to grasp an explanation one also needs to be clear on how the inferential relation between the explanandum and the explanans works. Both Khalifa and de Regt essentially make the same judgment in the case of the student who merely "knows by heart" the correct explanation—that she does not understand. The essential difference, however, is whether she lacks knowledge or skill. That is to say, for Khalifa this student lacks a crucial piece of knowledge about the inferential connection between the explanans and the explanandum. For de Regt on the other hand, grasping this inferential information is an acquired cognitive skill, separate from the information of the explanans itself.

We need not settle the dispute here, as for now it has been largely resolved by an academic interchange between de Regt and Khalifa in a recently published work (Lawler et al., 2023). There, de Regt acknowledges Khalifa's friendly effort to incorporate the essential use of skills in establishing the inferential relation between the explanans and the explanandum. Thus, it is safe to say that for both accounts the use of cognitive abilities and skills plays a central role for archiving the epistemic success of grasping. The flip side is that the lack of cognitive abilities and skills would thus lead to a lack of understanding. We follow this common ground and incorporate the lack of cognitive abilities and skills as one source of lack of understanding, illustrated in more detail in Sect. 3. Here we highlight that the lack of abilities and skills is not the only contributing factor for the negative dimension of understanding. As we shall see below, there are other factors (one of which we characterize as epistemic inclination in Sect. 2.4) that can lead to what we shall call misunderstanding, illustrated in more detail in Sect. 4.

To take stock, in this section we characterized the subjective components of understanding by combining the epistemic and psychological sides of grasping. In so doing, we defined subjectively grasping an explanation as:

(A) requiring abilities and skills to assess and evaluate how the explanation offers a solution to the explanatory seeking problem; and (B) having a positive phenomenal reinforcement of an insight of how the explanation solves the explanatory seeking problem.

Our next step is to combine the subjective and objective components of understanding, and provide a more explicit definition of explanatory understanding in Sect. 2.4. Based on that definition, we present the negative dimensions of understanding—lack of understanding in Sect. 3, and misunderstanding in Sect. 4.

2.4 Explanatory understanding and epistemic inclinations

In order to link the subjective and objective components of understanding, without having to develop a complete theory of explanations, an exhaustive account of grasping, and the role of abilities and skills, we found it useful to think about explanation as a type of problem-solving cognitive task.[8] This can be the analytic task of assessing and evaluating the information in an existing explanation, or arriving at a novel one.

In the former case, an explanation is a type of input. That input can be seen in some cases as requiring an analytic solution. For instance, when the cognitive agent is evaluating an explanation with a familiar structure and sufficient explicitness. In this case, grasping the explanation would be equivalent to analytically assessing and accepting it as correct. In other cases, this input can require a specific insight; for instance, when the explanation involves an unfamiliar, novel mode of reasoning, or is not sufficiently explicit. In this case, grasping the explanation would be equivalent to filling the informational gap within the explanation, or mastering the novel mode of reasoning exemplified by that explanation.

An explanation can also be a type of output. This occurs in cases where no previous explanation is available, and a novel one has to be developed. This again can be an analytic task, where the epistemic agent's previous knowledge is sufficient to solve a

[8] In fact, a theory that treats explanations as problem solving and connects understanding with insight has already been pioneered by Gary Hardcastle (n.d.). Whilst we find his manuscript as having a significant value for a potential research direction, the theory still has gaps. Its aim is to offer a general account of explanations (in terms of explicating the explanatory relation between explanans and explanandum), but it weighs too much on the subjective psychological dimension of explanations. Although we find the hypothesis in the manuscript highly intriguing, we leave its development for separate research. Since our present focus is to map the negative dimension of understanding, it is sufficient to specify that for us a theory of explanations as problem solving would concern the analysis of the cognitive task of discovering or evaluating the correct explanation, where the existing theories of explanations offer the criteria for relevance and information in such an explanation.

familiar problem. For instance, applying an explanatory pattern serially to cover phenomena within its scope (Grimm, 2010; Kitcher, 1989; Petkov, 2015). An explanation can also require a completely novel insight, in cases where a novel explanation must be discovered.

To generalize over these tasks, we can categorize them as requiring a gradual analytic solution, an insight, or both. As such, the objective component of the quality of the explanation, the employment of subjective cognitive abilities and skills, and the phenomenal hue of understanding would all be relevant for the resulting epistemic achievement. Given such framing, we find that relevant research from cognitive science and psychology bears on accurately linking the subjective and objective components of understanding.[9] As we shall see, this will also offer a previously unexplored dimension of understanding, as a process that forms epistemic inclinations.

To begin, both the literature on problem solving (Kaplan & Simon, 1990; Kounios & Beeman, 2014; Osuna-Mascaró & Auersperg, 2021) and on understanding (Baumberger et al., 2017) emphasize on the presence of some kind of phenomenology. This can be a form of insight, or a gradual increase of satisfaction from an analysis. Insight is seen as a sudden change in the type of knowledge representation or in the formation of a new concept relevant to the solution of a problem (Kounios & Beeman, 2014; Osuna-Mascaró & Auersperg, 2021). This sudden change is phenomenally accompanied by the "aha!" excitement (Kaplan & Simon, 1990; Kounios & Beeman, 2014; Osuna-Mascaró & Auersperg, 2021). Analytic solutions in turn are seen as the solver's deliberate manipulation of the "problem elements to discover the solution" (van Steenburgh et al., 2012, p. 475). This in turn is accompanied by a gradual, incremental warm feeling of satisfaction (Kounios & Beeman, 2014).

Of course, much the same way implicit learning can occur without phenomenal awareness, it is possible that understanding can happen without the supplemented meta-cognitive feeling (Cleeremans et al., 1998; Trout, 2002). However, we believe that explanatory understanding might be an exception to this rule. This is because explanations are conceptually explicit solutions to cognitive problems (i.e., explanatory seeking ones). Such solutions can be discovered (requiring an insight "aha!" moment) or analyzed (requiring deliberate analysis and a gradually increasing satisfaction of arriving at the correct solution). This makes arriving at explanatory understanding a more clear-cut epistemic effort. Consequently, the conceptual or discursive artifact of an explanation has to be in some sense explicitly present and complete. This artifact, as we noted (see Sect. 2.2), has to fulfill a given number of structural, informational, and ontological requirements, in order to solve the explanatory seeking problem. For such a completeness to be asserted, the epistemic agent has to have some kind of "break point" mechanism which would indicate that the construct satisfies the required structural, informational, and ontological properties imposed by the explanatory seeking problem. This "break point" signifies that the epistemic agent does not need to engage with the explanatory seeking problem anymore.

[9] It is worth mentioning that similar accounts that synthesize philosophy of science, epistemology, and cognitive science have already been pioneered by Khalifa et al. (2022). However, since this approach is relatively novel, and both the philosophical debates about understanding and cognitive science of insight and problem solving are all fast-developing research programs, we would be cautious to build our theory only upon on established claims.

Following the research of Ylikoski (2009) and Keil (2003), we believe that the phenomenal dimension of insight of understanding plays exactly this role. According to them, the sense of insight plays an important cognitive role, because it facilitates *cognitive economy*. Since we are limited epistemic agents, the pleasurable feeling of satisfaction that one reaches from a sense of insight, indicates that an epistemic goal has been achieved. Thus, such a feeling acts as a breakpoint, by signifying that the subject has received/or reached/sufficient information and no longer needs to cognitively engage with the problem. Similarly, an analytic solution can vary in its explicitness and depth, and consequently also requires some kind of cognitive mechanism that would indicate that a solution has been reached.

This role of a breaking mechanism (we argue) is played by a phenomenal sense of insight, or gradual satisfaction from a solution. Moreover, the relation between the positive phenomenology and epistemic economy, implies another important finding in cognitive science of problem solving—insight is more easily reached when one solves problems that call for familiar solutions (Lovett & Anderson, 1996; Schultz & Searleman, 2002). This preference for familiarity manifests itself also as a tendency to apply familiar solutions to novel problems (Lovett & Anderson, 1996; Schultz & Searleman, 2002). This in turn could lead to positive or negative epistemic outcomes. Crucially and in similarity to research in epistemology, a sense of insight can be misleading.[10] For instance, recent studies of Ammalainen and Moroshkina (2021) and Grimmer et al. (2022) have shown that triggering a false insight is possible by semantic priming or misleading hints. Some (Grimmer et al., 2022; Salvi et al., 2015, 2020) have interpreted these results as indicating that insight phenomenology functions to draw cognitive resources (attention) to ideas that are most consistent with one's implicit knowledge. Therefore, if past knowledge contains false beliefs, the resulting insight could be incorrect as well (Laukkonen et al., 2020; Webb et al., 2019). This interpretation of insight is also consistent with the paradigm which interprets insight as an implicit process of reasoning (Bowden, 1997; Grant & Spivey, 2003; Grimmer et al., 2022; Hattori et al., 2013; Laukkonen et al., 2018; Salvi & Bowden, 2020; Salvi et al., 2015; Schunn & Dunbar, 1996; Sio & Ormerod, 2009).

What this suggests to us is that, the sense of insight functions as an indicator of conformity between relevant, previous knowledge and novel knowledge. More narrowly for explanatory understanding, this suggests that the sense of insight can also function similarly. It will bias the epistemic agent to assert an explanation that conforms more easily with previously accepted explanations. The function of the sense of insight then is, to indicate that the explanation can be asserted, based on its conformity with the structural, informational, and ontological properties, that other already familiar explanations display. Obviously, such previously acquired explanations might vary in quality and thus the sense of insight would undoubtedly be judged epistemically, as a fallible indicator of understanding (Grimm, 2009). Again the emphasis is that, the sense of insight on its own is not to be blamed as being misleading, but that the implicit information that triggers it has epistemic flaws.[11]

[10] See the debate between Trout (2002) and Grimm (2009) on the reliability and the role of a sense of insight in understanding.

[11] Concerning analytic solutions more narrowly, as for instance the cases where the agent has to analytically evaluate an explanation that takes the form of deduction, Catarina Dutilh Novaes (2020) argues that deductive

However, for us it is more important to establish that, there is a tendency in epistemic agents to apply or receive insights from familiar solutions. And thus, they are more inclined to use such solutions to solve novel problems. This tendency to seek or assert explanations that display structural, informational, or ontological properties that are familiar or embodied by previously successful explanations, we call *epistemic inclinations*.

To develop this idea further, we establish that on the subjective side explanatory understanding requires:

(1) the relevant cognitive abilities and skills necessary to analyze or reach an insight to the correct explanation;
(2) this cognitive process is supplemented by a specific phenomenal hue.
To this we add that:
(3) the epistemic success of understanding also brings about the tendency of the habitual employment of previously proven to be successful, explanatory strategies.

We further hypothesize that the cognitive and psychological basis of such inclinations lie in the notion of mental sets which is closely related to problem solving (Schultz & Searleman, 2002). Mental sets are the tendencies of cognitive agents to apply previously successful solutions to problems that seem familiar (Lovett & Anderson, 1996). Mental sets are typically related to impairments of successful problem-solving tasks (Schultz & Searleman, 2002). This is because they lead to ignoring possible but unfamiliar solutions, in favor of familiar ones. A good illustration is an experiment of chess problem-solving, carried out by Bilalić et al. (2008). In the experiment, chess players were required to find a checkmate position with the fewest number of moves. Players were given a problem that has two possible solutions: one with five moves and the other only three moves. If the players were familiarized with the five-move solution, they would tend to use it, due to familiarity, instead of opting for the optimal but unfamiliar three-move solution.

Of course, solving scientific problems does not appear as clean-cut as checkmate scenarios which present definitive problems and solutions. As we mentioned, in scientific contexts explanatory seeking problems can be open-ended. However (in Sect. 2.2), we suggested that once certain structural, informational, and ontological requirements are imposed on a scientific problem, that problem can receive a relatively narrow set of correct solutions. We suggest that mental sets play a part in the formation of such narrower interpretations of scientific problems. The reason is that they involve a specific way of looking at the explanatory problem itself and trying to solve it using previously familiar strategies.

To further unpack the notion of epistemic inclination, we define it as: (1) a way to frame the explanatory problem based on familiarity, which leads to (2) asserting explanations (by analysis or discovery) that conform to specific structural, informational, and ontological requirements.

Footnote 11 continued
reasoning has a dialogical root and its properties (such as explicitness, self-evidence, etc.) are not completely innate. Thus, they are susceptible to variations among epistemic agents, based on their previous proficiency and experience in drawing deductions that succeed in convincing a relevant audience.

To give a simple illustration, if one has *an inclination to seek causal understanding*, when confronted with an explanatory problem, she would tend to (1) frame the problem as a causal one, and (2) try to provide a causal explanation that conforms to her implicitly/or explicitly/accepted theory of causality. This in turn would also postulate certain structural, informational, and ontological constraints on the evaluation and acceptance of potential explanatory solutions.

We do not claim that epistemic inclinations are completely equivalent to mental sets, much less to mental rigidity. In our case, we describe mental inclinations as the tendency of applying previously successful explanatory strategies to novel problems, and as such potentially facilitating or hindering understanding. This idea is also in tune with the notion of grasping, which captures the epistemic success of understanding as the ability to employ the explanation to solve similar problems (see Sect. 2.3 above). To add we will suggest that epistemic inclinations are subject to change, due to deliberate rational belief revision (in Sect. 6).

For now, it is important to emphasize that such inclinations can lead to a positive or negative epistemic outcome. It is negative in the sense that an agent might reject alternative explanatory strategies, exactly because they do not trigger a positive phenomenology. A phenomenology in turn depends on one's implicit epistemic inclinations. Consequently perhaps, the notion of epistemic inclination merits its own specific study, but here our goal is to provide a bedrock analysis of both understanding as forming such inclinations and their role in reaching understanding, lack of understanding, or misunderstanding.

With this we can expand our basic definition of understanding to:

> The cognitive achievement that occurs from the subjective grasping or construction of a correct explanation, that requires specific cognitive skills and abilities, is often accompanied by positive phenomenology, and leads to the development of relevant epistemic inclinations.

With this definition in mind, we can examine the epistemic phenomenon of understanding as being a consequence of an alignment[12]—an alignment that occurs between the subjective and objective components of understanding. This alignment entails (1) the right abilities and skills of an epistemic agent, plus the positive reinforcement role of the phenomenal sense of satisfaction or insight, and (2) the correctness of the explanation itself.

In the context of explanatory understanding, the alignment of its components leads to the agent solving the explanatory problem to her satisfaction and thus to the epistemic success of understanding. Conversely, misalignment leads to the problem remaining unsolved. This can occur in two different situations. Firstly, when any of the components is not in order, such as on the subjective side when the agent lacks abilities and skills, or on the objective side when she grasps an incorrect explanation. Or secondly when there is a misalignment between the components themselves, in cases where a correct explanation engages an agent who in principle has sufficient abilities, skills, and knowledge to grasp an explanation, but rejects it because it does

[12] The usage of the word "alignment" here has a functional connotation, in that it denotes bringing the parts into proper arrangement to allow the whole to function coordinately towards a certain end.

not conform to her epistemic inclinations. The former case, as we shall see shortly, leads to a lack of understanding, whilst the latter to misunderstanding.

3 Lack of understanding

To shift our attention to the negative dimension of understanding, we first characterize cases of lack of understanding, as situations where people fail to solve the explanatory seeking problem they have set. For us, all cases of lack of understanding are cases where no correct explanation is constructed or grasped. This implies that a lack of understanding can occur when:

(1) *subjectively the agent lacks*[13] *the cognitive abilities and skills* necessary to grasp or construct a correct explanation; *or*
(2) *the agent grasps an objectively incorrect explanation*—an explanation that fails some of the three requirements outlined in Sect. 2.2.

On the subjective side, when one's cognitive abilities and skills are inadequate to grasp or construct a correct explanation, one lacks understanding. These cases are widely discussed in the literature, typically by the skill-based accounts. In Sect. 2.3, we illustrated this type of epistemic failure by the example of a student who lacks the reasoning skills and simply memorizes the explanations in the textbook. Relative to the phenomenology of understanding, the positive feeling of understanding does not occur in cases that fall under this category.

It is also possible that the agent grasps an incorrect explanation. Such an explanation can fail any of the three requirements for correct explanations: structural, informational, or ontological.

Starting with explanations that are structurally incorrect, as we mentioned such explanations can fail some of the structural requirements of the relevant theory of explanations. For instance, an agent can grasp a mere valid deductive explanation that fails to exemplify a lawful generalization; or a causal conditional that is based on spurious causes.

When the informational requirements are violated, we encounter a more complicated scenario. This is because the criterion which determines if an explanation is informatively adequate is contextually sensitive. It means that the satisfaction of the informational requirements of an explanation depends on the specific requirements of the explanatory seeking problem itself. We tried to provide such a contextually sensitive criterion by defining explanations as informationally *sufficient*, *inadequate*, *incomplete,* or *excessive*. Importantly, as we showed in Sect. 2.2. the same explanation can turn either way by simply switching or redefining the requirements of the explanatory seeking question. Having in mind such contextual sensitivity, we can nevertheless generalize about the understanding that such explanations will provide.

[13] Here we must note that human abilities and skills are limited and fallible. Therefore, this failure can cover a broad spectrum of phenomena, from cases where such abilities and skills are simply not present or developed, to cases where such abilities and skills in principle cannot be developed (for instance where the explanatory seeking problem exceeds human reasoning abilities). This also includes special cases where such cognitive abilities and skills are faulty, for instance one misremembers, or fails in calculating, or seeing an error in a logical argument, etc.

Obviously sufficiently informative explanations generate a solution to an explanatory seeing problem and as such count as correct, and thus providing understanding. The negative cases however are more complicated.

Whenever an explanation is *informatively inadequate*, it wholly misses the mark of the informational requirements of the explanatory seeking problem, and as such does not generate understanding. A myriad case of transitivity of causality in the literature of causal explanations can serve as a good example. Reiss (2015), Northcott (2008), and Hitchcock (2003), for instance discuss the following narrative: "Captain yells 'fire', trainee fires. Upon hearing the command, victim ducks. The bullet misses the victim who survives unscathed." For the question: "What saved the victim's life?", one answer could be: "The victim survived, because the captain yelled fire." The literature on causal explanations converges on the idea that such an explanation misses the mark of informativeness and as such leads to a lack of understanding.[14]

An even more complicated scenario occurs when explanations are *incompletely* or *excessively* informative. Here we defend the claim that an incomplete explanation provides some explanatory information, but it does not lead to an increase of understanding. This is because the explanation fails to meet the exact requirements of the explanatory seeking problem. However such an explanation, if taken regarding a different but related explanatory seeking problem, can be seen as satisfactory, and therefore as increasing one's understanding. For instance, if the explanatory seeking problem a scientist has set is to understand the exact mechanism of long Covid complications, she can fail the task with an incomplete explanation that "Covid19 causes long Covid". This explanation is incomplete, because the scientist is at least clear that Covid19 is the main cause of long Covid. Hence, the offered explanation does not increase her understanding. For, she has failed to solve the problem of describing the mechanism of long Covid. However, for the question: "What causes such a condition?" the answer "Covid19 causes long Covid" would be sufficiently informative and increase her understanding.[15]

Finally, *excessively informative* explanations present an even more curious case. On the one hand, it is possible that the explanation seeker can manage to simply "cut off" the informational content at the right place, and as such reduce the excessive

[14] Again, we emphasize that the informational requirement must be taken narrowly, as relevant to the specific requirements of the explanatory seeking problem, and the theory of explanations that provides a solution. Consequently the explanation: "Because the victim ducked." would perhaps count as sufficiently informative and correct. However, an addition with a regressive question "Why did the victim duck" would have to include the fact that the captain gave a clear signal that the victim could read. Similar claims are made by Reiss (2015) who develops the idea that causal explanations are contextually dependent. Others (Potochnik, 2017) have followed with the same line of arguments.

[15] There are a myriad of ways in which the literature on understanding and explanations copes with these issues. For instance, Potochnik (2017) claims that a successful causal explanation is determined in part by the exact interest of the explanation seeker. As such whether the same causal explanation is epistemically successful, depends on if it satisfies the interest of the explanation seeker. Khalifa (2017) on the other hand defines understanding as a notion relevant to the explanandum taken as a topic. As such the understanding of the explanandum depends on the amount of available explanatory information about it. In other words, the understanding depends on what he calls an explanatory nexus. Such a nexus can grow every time new explanation is added, but it does not grow when one fails to solve a new explanatory seeking problem, even though prior explanatory information might be present.

explanation to a sufficiently informative one. Again provided with a detailed explanation of the mechanism of long Covid, one can manage to extract information for the reduced explanation “Covid19 causes long Covid”, for her problem “What causes long Covid?”. On the other hand, the added complexity of such explanations might make it difficult to understand the explanation itself, and as such a failure to understand the explanation would also lead to a lack of understanding. Using the same example, one might not manage to extract the explanatory information needed to solve the problem.

Moving to the ontological requirements for an explanation, the situation seems to be, at least on the surface level, similar to that of structural requirements. If the ontological requirements for an explanation are not satisfied, that explanation consequently would not generate understanding. However, in scientific contexts the informational and ontological requirements can pull in different directions. For instance, if one’s ontological requirements are strictly factive but the informational requirements are less stringent, one is often left with a choice between asserting a generalization that is strictly speaking factively false for the explanandum, and seeking to provide an overly complicated explanation. The debates between quasi-factivists and factivists, along with the studies of the role of idealization and modeling in science, provide abundant illustrations. However, we must note that the debates about both the factivity of explanations and the role of idealization in modeling take a significant portion of the contemporary literature in philosophy of science, and therefore we cannot provide a serious, systematic study here.

However, our intuition is that in many scientific scenarios, the complexity of jointly satisfying both the informational and ontological requirements for an explanation can present a challenge. Therefore, some trade-offs are often made. We shall return to these issues later (see the Appendix), but for now we can provide a simple example. The informational requirement for solving a causally complex problem might be such that, if one loosens her ontological requirements, the problem might be solvable. For instance, Potochnik (2017) has suggested that this is the case when one asserts, what she calls, causal-pattern explanations. Such explanations describe causal relations, but also assume a level of idealization. Nevertheless, Potochnik notes that these explanations generate understanding.

On the other hand, if one asserts a rigid factive ontological requirement, it will weigh on the information of a potential causal explanation to such an extent that a complete factive causal explanation might not be achievable for human agents.

With these considerations in mind, our goal here is to explore a hitherto unrecognized possibility—situations where the epistemic agent’s skills or abilities are sufficient for the task of grasping what can be asserted as a correct explanation, but she rejects that explanation nevertheless. We postulate that this rejection can occur, because the agent’s epistemic inclinations are incompatible with such an explanation. In such a case the agent might seek an alternative that fits better her epistemic inclinations. We categorize such and related cases as situations of misunderstanding.

4 Misunderstanding

Misunderstanding has two common usages in the English language: *failures to understand correctly*, and *disagreements between two parties*. We find these lexicographical definitions useful, and our use of the word "misunderstanding" in this paper is populated by them. However, we must clarify the first usage ("failures to understand correctly"). As noted above, we are inclined to delineate between "lack of understanding" and "misunderstanding", at the point where a lack of understanding concerns epistemic failures that result in or from an incorrect explanation. Misunderstanding, on the other hand, is the rejection of a correct explanation, in the presence of the cognitive abilities and skills to grasp it. In this sense, misunderstanding can be seen as a negative propositional attitude the agent has towards a correct explanation. Consequently, the epistemic outcome from such a propositional attitude can be seen narrowly as negative (i.e., as a lack of understanding), because when one misunderstands the explanation, she does not understand the explanandum through that explanation.[16]

This rejection can occur whenever there is a misalignment between some of the structural, informational, or ontological features of the explanation and the inclinations of the epistemic agent for them. As we mentioned (see Sect. 2.4), epistemic inclinations are tendencies to assert or apply familiar explanatory strategies. These strategies might have been successful in solving previous explanatory seeking problems, but that does not necessarily mean that they will always be applicable to novel ones. Consequently, and in similarity with the suggestions from the literature on problem solving, explanatory problems which require unfamiliar solutions would be typically harder to solve. We suggest that this difficulty might be due to the fact that the epistemic agent's inclinations are ill-suited to address the specific explanatory problem.

A further negative outcome might occur in cases where there is a misalignment between the features of an otherwise correct explanation, that the epistemic agent considers, and her inclinations. Consequently, even though the agent has the cognitive abilities and skills to evaluate the explanation, as a correct solution of a specific explanatory seeking problem, the positive outcome of asserting the explanation might be blocked by a lack of positive insight. We hypothesize that this lack of positive phenomenology plays an important contributive role for misunderstanding. This is because the phenomenology of understanding or insight and the formation of epistemic attitudes towards an explanation are intimately linked. As we mentioned, the literature on problem solving and insight seem to suggest that the phenomenology of insight is a result of an implicit cognitive process of evaluation of a solution to the problem, in light of relevant previous experiences. This evaluation is suggested to be based on conformity (Grimmer et al., 2022). Concerning the narrow case of explanatory seeking problems, this suggests that the implicit evaluation of the explanation

[16] To frame this as a problem-solving task, when one rejects an otherwise correct explanation not because a lack of abilities or skills to grasp it, one still fails the specific task of understanding the explanandum through the specific explanatory seeking problem and its solution postulated by that explanation. According to our definition, this will lead to lack of understanding. However, one fails the task so to speak in a "special way". That is only because one's insight has not been triggered. An epistemic agent can give up the task, or seek further insight from an explanatory alternative that satisfies her inclinations. Nevertheless, that alternative explanation would concern reframing the explanatory seeking problem differently and in turn can result in another line of understanding of the explanandum.

is based on such conformity with familiar types of explanations and their structural, informational, or ontological properties. This evaluation is often implicit, as the literature suggests (Grimmer et al., 2022; Laukkonen et al., 2021). Consequently, only the outcome, as the presence of a positive feeling of an insight from the given solution (an explanatory one in our case) and the tendency for its acceptance, is available to the epistemic subjects. This leads us to believe that the feeling of insight can be seen as reinforcing one's attitude towards an explanation, even when we are engaged with an explicit evaluation of the features of the explanation. Having these ideas in mind, we can provide a definition of misunderstanding as:

> Misunderstanding is the epistemic failure, that occurs when a correct explanation is offered, that explanation can be in principle grasped (as the epistemic abilities and skills of the explanation seeker are sufficient for its grasping), but understanding of the explanandum through that explanation is blocked by a mismatch between the features of the explanation and the epistemic inclinations of the explanation seeker.

To illustrate this definition, we suggest a few rough examples of misunderstanding. These examples show a mismatch between the properties of the explanation and the inclinations of the explanation seeker.

For instance, Jansson (2014) argues that the debate between Leibniz and Newton, is partially a debate about the correct theory of causation and if explanations of gravitational phenomena, as provided by Newton, are instances of causal explanations. According to Jansson, this historical debate has a modern twist, because it is investigated differently by two contestants for what would count as the correct theory of causal explanations—those of Woodward (2003) and Strevens (2008). Arguably then both theories explicate and develop different perspectives about what causality is, and form a different inclination about the structural, informational, or ontological properties of causal explanations. We suggest that, this line of dispute is possible exactly because each side addresses causal problems, by interpreting them as structural explanations within their respective theory, and thus providing informational and ontological requirements for what would count as a correct explanation. This results in rejecting their opponent's explanations as structurally incorrect, insufficiently informative, or ontologically lacking. We propose that such a genuine scientific and philosophical dispute would be solvable, if both sides could find convergence on the structural, informational, or ontological properties of causal explanations.

We suggest that a similar scenario occurs within the dispute between proponents of distinctively mathematical explanations and causal explanations. An inclination for a deductive solution to an explanatory problem might lead to the rejection of a causal solution, or vice versa. This is because the two sides differ in their inclinations towards specific explanatory structures[17] and their informativeness. As we shall see in the case study (Sect. 5), the proponents of each kind of explanations claim that their rivals have a misunderstanding (as we define it) of what the correct explanation would be. To use our terminology, this misunderstanding concerns a case where proponents of distinctively mathematical explanations claim that their rivals misunderstand

[17] Deductive or causal ones.

and assert an excessively informative explanation. Proponents of causal explanations, similarly, claim that their rivals misunderstand and assert an insufficiently informative explanation.

Finally, people might diverge on their requirements for the ontological properties of an explanation as well. Some might identify the truth makers with factual correspondence, whilst others might embrace factivly weaker, pragmatic considerations. This in turn, we suggest, can lead to a disagreement over the role and level of idealizations that are acceptable in explanations, derived from idealized models (see the Appendix).

In all these cases, however it is important to note that the rejection of what can potentially be a correct explanation is not due to a lack of abilities or skills to grasp it. In all these cases, the epistemic agents involved are trained and respected scientists or philosophers for whom it can hardly be claimed that they lack specific abilities or skills to evaluate the explanations in question. Instead, we claim that epistemic inclinations, the intuitions they create, and the subsequent biases in favor of specific explanatory strategies all play a major role in such disagreements.

The picture we have painted so far might seem to lead to a perspective-dependent and perhaps relativistic account of understanding. However, very often participants in such disputes share a common scientific goal and consequently, convergences are found on some structural, informational, or ontological requirements for asserting one type of explanatory solutions over others. As such the logical outcomes of such disputes can be of two kinds:

(1) *Pluralistic* where: the debaters can agree that both explanatory strategies provide correct explanations but relative to different structural, informational, and/or ontological framings of the explanatory seeking problem. During the disputes, both parties were misunderstanding each other's otherwise correct explanations.
(2) *Monistic* where: the debaters can agree that one of the strategies has been misguided, and the other is the singly correct explanatory strategy. During the disputes, the party in error was misunderstanding their rival's correct explanation.

To both outcomes the role of epistemic inclinations is similar to that of mental sets. They would facilitate understanding through familiar explanations and hinder epistemic success in the face of unfamiliar ones. We hypothesize that both the role of mental sets in hindering or facilitating insights, and the need for cognitive economy are relevant for such outcomes. Similar to the literature on insights and mental sets, we postulate that this could be both a positive and a negative phenomenon. Negative in a sense that it could lead to a disagreement between scientists on two otherwise genuine explanatory strategies, or to make one inclined to reject an otherwise correct explanation. But positive in the sense that it could lead to the discovery of complementing explanations, or even to a broad synthesis between what were seen as initially incompatible explanatory strategies (see the Appendix).

5 Causal VS distinctively mathematical explanations: a case study

The debate on distinctly mathematical explanations concerns whether explanatory understanding of some ontic facts can be achieved through purely mathematical framing of these facts, without implicit or explicit information on the causal genesis of the represented facts. Proponents of distinctively mathematical explanations claim that, in special cases where the representative mathematical structures constrain the phenomena, information about such mathematical structures can provide understanding for the represented facts (Lange, 2013, 2017). Proponents of causal or broadly ontic accounts of explanations reject this possibility (Bangu, 2021; Craver & Povich, 2017). Instead, they claim that an explanation would be in important ways incomplete without the presence of relevant causal or broadly ontic information.

Albeit ongoing, this debate in philosophy of science is of significant interest to us. This is because the two sides argue about essentially the same factual explananda, but disagree on which type of explanations provides understanding for these explananda. Moreover, it can hardly be claimed that any side lacks the skills or abilities relevant to grasp the opponent's arguments and explanations. Therefore, we propose that the debate can be conceptualized as an interesting case of misunderstanding where epistemic inclinations might be partially responsible for the ongoing disagreement.

Our central claim is that the two parties in the debate have different epistemic inclinations (to seek mathematical or causal understanding respectively). As we postulated in the previous section, such a disagreement may have two potential outcomes: *pluralistic* or *monistic*. In case of a *pluralistic outcome*, the two parties would agree that both their explanatory strategies are correct. However, during the debate when the parties were actively disagreeing with each other, they also misunderstood the other's *otherwise correct explanations.* If the latter *monistic outcome* occurs, a single correct explanation would be agreed upon. In this case one of the sides would concede that its explanatory strategy was inadequate for solving the problem. Thus, during the debate the party in error was misunderstanding their rival's correct explanation.

Before presenting the debate itself, we make some clarifications on our motivation and position. Firstly, our aim here is not to present a knockdown argument for any side. Instead, we analyze the debate as it stands—an ongoing debate within philosophy of science. Our motivation to choose this debate, as an illustration of our framework, is exactly due to its present state of being undecided. This essentially permits us to use it as an illustration of the two possible outcomes we described above.[18] With this we intend our analysis to serve as an illustration of a general methodology—an illustration of how our account of misunderstanding can be applied to clarify positions in tenacious disagreements that can conclude with any of these logical possibilities.

Analyzing the debate from the perspective of pluralism might at first seem like a more natural solution. However, we shall see that a deeper discussion on the *ontological requirements* for both kinds of explanations can lead us to assert that, the monistic reading is a viable option as well.

[18] This is dictated by economy, because if we were to choose a narrower case, we would have to provide two examples: one of a debate with a pluralistic conclusion and one with a monistic outcome.

To begin, the canonical case around which the discussions evolve is: *A parent who has twenty-three strawberries and three children cannot distribute the strawberries evenly without cutting any (strawberries!)* (Lange, 2013).

Here the proponents of distinctively mathematical explanations claim that a correct explanation would be: "*Because 23 is non-divisible evenly by three and this mathematical fact constrains the possibility of dividing the strawberries evenly*" (such as Lange, 2013). While the proponents of causal explanations usually propose that a correct explanation would have to be: "*Because the physical system is causally constrained in such a way that both the facts of having 23 strawberries and 3 children hold, the fact that 23 is non-divisible evenly by three explains the failure of the parent.*" (such as Bangu, 2021).[19]

If one frames the problem in two ways, it is easy to see how both parties can be right:

"Why does the parent fail *to divide 23 by 3* (strawberries) *evenly*?" and "Why does the parent fail *to physically divide* 23 by 3 *strawberries evenly*?". These two different framings can be further analyzed as a case of diverging epistemic inclinations. As defined in Sect. 2.4, an epistemic inclination is: (1) a way to frame an explanatory problem based on familiarity, and (2) the tendency to assert explanations (by analysis or discovery) that conform to familiar structural, informational, and ontological requirements.

The first one shows that a mathematically inclined explanation seeker, perhaps due to her familiarity and training in solving mathematical or purely theoretical problems, would tend to frame the facts of the explanandum in a formal context. The second posit entails that subjects would also seek an answer that is conceivable within a narrow theoretic context. In such a context an explanation, similar to a mathematical proof, would be preferable. They would thus tend to assert an explanation that takes the form of a deduction and involves a mathematical proof. A relevant and telling quote from Lange's book which seems to suggest this idea is:

> "By the same token, "word problems" in mathematics textbooks are full of allusions to facts that have distinctively mathematical explanations…" (2017, p. 25).

On the other side of the debate, proponents of causal accounts might be inclined to address the same problem with a different cognitive habit. Undoubtedly causal reasoning is our most common method of coping with the surrounding environment. For instance, Woodward (2003) claims that some animals and human infants develop a concept of causality, because such a concept is important to predict what would happen if they interacted with the environment in a specific way. Similarly, ontic accounts of explanations usually claim that the central aim of scientific endeavors is to predict

[19] Bangu (2021) claims that the explanatory seeking problem can be represented more generally as a task of putting one medium-sized physical object into two spatially separated containers at the same time, such that each container holds an object identical to the initial object. The reason why this is impossible is that the fundamental physical law of conservation of mass holds. However, if the law of conservation of mass weren't in place, another mathematical fact could be applied which would lead to the parent successfully distributing the strawberries. Bangu claims that such a fact is the Banach-Tarski theorem. According to which one three-dimensional sphere can be decomposed into several disjoint subsets, that can be reassembled into two spheres identical to the original one.

and control the changes in our environment. This ensures the dominant role of causal explanations.

Framing the debate in this way shows that both parties have different epistemic inclinations. These inclinations underlie the way they formulate different explanatory seeking questions relative to the same explanandum. Following this line of reasoning also allows us to view each side as claiming that the other provides excessively or incompletely informative explanations.

From the perspective of mathematical explanations, adding further causal information leads to *excessively informative* explanations. Lange (2017) for instance seems to defend this line of argument. He claims that the addition of causal information[20] to a distinctively mathematical explanation is explanatorily inert and does not contribute to the understanding of the explanandum. On the other hand, a causal reading of the explanatory seeking question dictates that purely mathematical answers are *incompletely informative*. For instance, Bangu (2021) and Bueno and Colyvan (2011) claim that whilst mathematical truths have a contributive role to an explanation, an additional causal fact plays the actual explanatory role (e.g., the law of conservation of mass).

One can be tempted to read the debate under a pluralistic framework, especially given the fact that perhaps each party's intuitions are informed by equally important cognitive abilities (that to deduce and generalize; or that to orient oneself in a world of causes and effects). However, a deeper examination of the debate shows that the challenges from both sides are also about incompatible ontological commitments. That is to say, each party's inclinations also concern different requirements for the ontological properties of explanations. If one pursues this interpretation of the debate, perhaps one would reach a *monistic* conclusion.

For instance, Lange (2017, 2021) claims that explanations in physics are organized in a nested hierarchy of necessities where higher order mathematical facts have a priority by constraining ontic facts. This then, according to Lange, makes one committed to deeper claims about the ontological status of mathematics. Essentially his argument is that both nominalism and Platonism fail to account for distinctively mathematical explanations. Instead, such explanations are possible only if Aristotelian realism is true (Lange, 2021). According to it, mathematics concerns mathematical properties possessed by physical systems, and as such mathematical properties can constrain physical ones (Lange, 2021). This claim obviously touches on the ontological status of mathematical entities and is made even more pressing, because the argument of Lange (2017) is aimed at explaining the role of essential explanatory tools of physics such as the Lorentz transformation.

Such an ontological commitment to the status of mathematical facts cannot be jointly true with the opposite claim coming from causal accounts. For instance, Bueno and Colyvan (2011) defend an indexing account of the role of mathematics in science. According to it, mathematical objects are not existing entities. This mismatch is further aggravated by the fact that both ontologies are not compatible.[21]

[20] Causal information beyond the bare minimum to describe the explanandum as a stable casual system, that can satisfy the mathematical explanans.

[21] Mathematical facts cannot both exist in the Aristotelian sense and serve the role assigned to them by the indexing account, argues Lange (2021).

Given the mismatch between the two parties' inclinations about the ontological status of explanations, the solution can only be *a monistic* one.[22] One ontology about the status of mathematical objects and their role in science has to prevail. Consequently, the party in error has to consent that they have offered an ontologically incorrect explanation. They should also acknowledge that they have misunderstood their rivals correct explanations.

To conclude, here we would not develop further arguments that might tip the scale for any side of the debate, or for a pluralistic reading. Instead, this discussion should serve as an illustration of how our account can be applied to similar disputes about scientific explanations. As we showed, it could be used to analyze the outcomes where one side prevailed, or if they ended up in mutual recognition.

6 Conclusion. epistemic inclinations impasse or progress?

It might seem that we have painted a rather grim picture for solving tenacious debates in philosophy and science. However, one needs to recall that we developed the notion of epistemic inclination in relation to mental sets, but different than mental rigidity.

For us, epistemic inclinations result from receiving a positive phenomenology that occurs after a successful problem solving. As we mentioned, the cognitive science of insight suggests, that insight phenomenology functions to "select" ideas from the stream of consciousness by drawing attention to ideas that are most consistent with the agent's implicit knowledge (Salvi et al., 2015, 2020). This makes the available background knowledge essential for the development of inclinations. Importantly however, our background knowledge is often subject to rational belief revision. It is possible that the introduction of new arguments and ideas can result in a revision of one's implicit belief system. Therefore, as in any debate with the introduction of further arguments and sufficient open-mindedness from both parties, a consensus might be reached. This is because each party's inclinations about the structural, informational, and ontological features of what constitutes a correct explanation would become explicit. Given such explicitness, one can also revise or further affirm their epistemic inclinations in light of new information. Consequently, a consensus can be reached. Such a consensus would occur if the parties reach an agreement on the structural, informational, and ontological properties of the potential explanatory solutions to the scientific problem they seek to resolve. Therefore, we believe that scientific misunderstanding due to diverging inclinations could also lead to progress.

Acknowledgements This work is supported by The Fundamental Research Funds for the Central Universities, project number 1233300003. Many thanks to the audiences of the following conferences for providing valuable feedback on this paper or the content it covers: CSHPS 2021, Canadian Society for the History and Philosophy of Science Conference; EENPS 2022, East European Network for Philosophy of Science Fourth Conference; PSA 2022, the 28th Biennial Meeting of the Philosophy of Science Association. Special thanks to Philip Kitcher for a helpful commentary on misunderstanding. We wish to express gratitude to Dai

[22] This would entail a knockout argument offered by one side. However, a pluralist can also claim that both ontologies are tightly knitted to their respective arguments about explanations. Therefore, arguments about such ontologies cannot be settled in a theoretically neutral way. This line of argument is usually pursued by agnostics about the possibility of a solution of the "scientific realism VS instrumentalism" debate (Chakravartty, 2017; Forbes, 2017).

Haiqiang, Gennady McCracken, Lilia Gurova, and Mark McCullagh, for their critical comments and support of the present research. We would also like to thank three anonymous reviewers for their helpful comments which improved the manuscript significantly.

Declarations

Conflict of interest This work is supported by The Fundamental Research Funds for the Central Universities, project number 1233300003. The authors have no relevant non-financial interests to disclose.

Appendix

Understanding and its failures in modeling predator–prey interactions

The growing literature on modeling and idealization compels us to provide an appendix to discuss the theoretical implications and possible treatments of our account in this regard. In the broad context of modeling, idealization is often discussed as a feature or method of modeling (Batterman, 2009; Frigg & Hartmann, 2020; Rohwer & Rice, 2016). Hence the appendix will summarize several kinds of models, and discuss the role of idealization in modeling, in relation to our framing of explanatory understanding, lack of understanding, and misunderstanding.

To start with, some models are treated similarly to explanations, in that their explanatory power is measured against the standard set for explanations. For instance, Craver (2006) separates explanatory models from merely phenomenal models by the former's affordability of manipulation in the Woodwardian (2003) sense. Similarly, Potochnik (2017) claims that any representation that can reliably generate understanding is an explanation, including heavily idealized representations, and that seems to include all models that generate understanding. By the same token, we are inclined to hold similar positions and claim that there is a significant overlap between modeling and explanatory understanding. Nevertheless, it is worth mentioning that scientific endeavors in modeling cover a broad spectrum of activities, including graphics (de Regt, 2017), examplars (Elgin, 2017), or even model organisms such as the *Drosophila*, all of which can contribute or even provide direct forms of understanding. However, this broad spectrum of what is considered a model, and an idealization, significantly complicates making any non-surface level generalizations and analyses.

Therefore, to restrict the discussion we focus only on models and idealizations taken narrowly, as mathematical structures. The properties of such models can be clearly conceptualized. When these models are used to represent a given empirical system, one can also use them to derive propositional claims about the empirical target. Such models can be loosely investigated as sets of functions which interpret as their parameters some of the quantifiable features of the observable system. Given an input that substitutes the parameters with values derived from observations, these functions will generate conclusion-like statements about their empirical targets. If the model manages to be representative of its target, the conclusions made by the model, along with its assumptions, can be ordered into explanatory inferences (Petkov, 2019). For the sake of clarity, we consider models to function in such a way.

Here we also exclude discussions about cases where models and idealizations of various kinds can provide understanding directly, without generating any kind of conceptually framed explanations. We concede that this is certainly possible for material models, simulations, exemplars, etc. (Frigg & Hartmann, 2020). Similarly, we will not discuss cases where models do not directly relate to understanding but merely contribute to the construction of explanations (Rohwer & Rice, 2016).

Focusing only on the role of idealization in modeling, does not provide a unified picture, either. Idealization, as a modeling technique, serves a variety of purposes. Ontologically, it can simplify ("Aristotelian idealizations", Weisberg, 2007) or distort facts ("Galilean idealizations", McMullin, 1985), while epistemically it can be used to isolate the desired explanatory patterns or structures (Potochnik, 2017). Nevertheless, we can narrowly investigate idealizations as deviations from the facts. As for the reason for such deviations, we agree with Batterman (2009) and Potochnik (2017), that idealizations reduce the complexity of empirical facts, thus making the target system easily understandable by limited cognitive agents such as us.

With these clarifications in mind, we focus on a case where an idealized mathematical model plays an explanatory function. The resulting explanations then, as we shall see, can lead to understanding, lack of understanding, and misunderstanding.

We suggest that misunderstanding can occur due to diverging inclinations about the informational and ontological properties of the explanations derived from such models. This tension can again lead to scientific disputes, where different parties do not lack the abilities and skills to grasp their rival's explanations, but still misunderstand (in the sense we developed) each other.

To illustrate this idea, we shall focus on the so-called functional response debate in ecology (Abrams & Ginzburg, 2000). This debate concerns two types of idealized ecological models of predator–prey interactions, that interpret the so-called functional response differently. To clarify, in these predator–prey models the functional response is a function which governs the intake rate of the predator in relation to food density.

In the classical Lotka-Volterra (1926) models, the functional response depends only on the density of the available prey. Various one-step de-idealizations to this functional response were initially made. For instance, type II functional response, which is a function that describes the decreasing intake of predators because of satiation (Holling, 1959). Such functional responses were generalized under the concept of *prey-dependent functional response*.

Rosenzweig (1971) used a similar prey-dependent functional response to show the highly controversial conclusion that if the prey-carrying capacity of a stable predator–prey system is increased sufficiently, both the predator and prey populations would be driven closer to extinction. This outcome became known as *the paradox of enrichment,* and had attracted numerous empirical and theoretical studies. Direct evidence for the occurrence of the paradox, however, remained largely confined to microbiological experiments (Arditi & Ginzburg, 2012, p. 120).

Arditi and Ginzburg (1989) suggested another functional response, the so-called *ratio-dependent functional response*, represented as a ratio between the densities of the predators and prey. The justification of defining the functional response as a ratio rests on the idea that predators will interfere with each other for the available prey, directly or indirectly. Importantly the resulting equations do not exhibit the counter-intuitive effect of destabilization due to enrichment. Nevertheless, this model again is not without flaws. It predicts that predators can survive over an arbitrary small-prey population, and it also generates incorrect results about food chains at low levels of enrichment (Ginzburg & Jensen, 2008).

Importantly both models are highly idealized. This means that a factive ontological evaluation of the explanations, derived from such models, must evaluate them as false. Nevertheless, these models have an overlap in their predictions (standard cases of equilibrium between predators and prey), but diverge in their predictions in extreme cases of high or low enrichment. Given the generality of both models, a dispute arose about which model could generate understanding (through its descriptions of predatory-prey dynamics) for which kind of predator–prey ecologies.

To frame this debate in our terms, both models were used as explanatory since they provided sufficiently accurate descriptions of the general features of predator–prey dynamics. Consequently, their explanations could be read as approximately true (Petkov, 2019) or having a pragmatical value based on their relative predictive success.[23] In other words, both models could be considered explanatory, since they did describe the general behavior of predator–prey populations and, as such, can be taken as exemplifying what Potochnik (2017) calls causal-pattern explanations. These models can be seen as jointly sufficiently informative, for the most general cases of understanding the ecology of predation, as they correctly describe standard cases of stable equilibrium between predators and prey. However, on a closer look upon closer examination, the explanations derived from these models yield contradictory results regarding what would happen if enrichment occurred. Thus, explanatory derivations from both models, if focused narrowly on the same population dynamics, cannot be jointly correct. Hence a debate about their applicability ensued.

In our framework, this means that each model, if taken to represent an agreeable explanatory structure (a causal-pattern explanation), generates *sufficiently informative conclusions* for standard population dynamics, but *inconsistent* and jointly contradictory conclusions for the extreme cases of extinction, due to enrichment or impoverishment of the environment. One must also assert that each model generates relevantly accurate descriptions of predator–prey populations in standard cases of equilibrium, and hence generates understanding. But they also generate inconsistent and jointly incorrect conclusions about extinction scenarios of the very same population, and hence result in lack of understanding.

[23] As we mentioned above, such models generate conclusions about their targets' behavior that can be taken as both explanations and predictions. This is because predator–prey equations are typically simple differential equations.

This tension between inconsistent, equally idealized models and their informativeness was at the center of the debate. Moreover, our study of the debate (Petkov, 2019) and its tone (Abrams & Ginzburg, 2000) lead us to postulate that the parties involved defended their preferred model, based on the *inclinations* they had developed over years of conducting research with these models.

Here we can frame the debate as a case of *misunderstanding*, where theoretic ecologists of equal merit were rejecting the understanding their rival's models generated, even though both sides defended equally idealized models.[24]

Without going into details here, radical solutions for abandoning one of the models were proposed (Abrams & Ginzburg, 2000). Nevertheless, a consensus was eventually reached (Abrams & Ginzburg, 2000). This consensus concluded that each model could be confined to a specific context of applicability. Years after reaching the consensus, a synthesized model that subsumed the previously conflicting models into a consistent structure was also developed (Ginzburg & Jensen, 2008).

Our analysis of the historical conclusion of the debate suggests that understanding was reached, when each party overcame their inclinations, considered their rival's model's applicability, and realized where their preferred model was at fault (Petkov, 2019). Therefore, this case of explanatory understanding based on idealized models, falls under the pluralistic outcome mentioned in Sect. 4. Importantly, as mentioned earlier, the conclusion of the debate also led to the development of novel approaches to modeling and hence understanding predator–prey dynamics (Ginzburg & Jensen, 2008).

Regrettably the limited space of this study confines us to capture this debate only roughly within our framework. However, we hope it shows that our conceptual framework is applicable to cases of understanding and its failures for modeling and idealization. Here we cannot provide an exhaustive illustration of how our account can cover debates about the understanding derived from idealized models. Such a project will hopefully be carried out in the near future.

References

Abrams, P. A., & Ginzburg, L. R. (2000). The nature of predation: Prey dependent, ratio dependent or neither? *Trends in Ecology and Evolution, 15*(8), pp. 337–341. https://doi.org/10.1016/S0169-5347(00)01908-X

Ammalainen, A., & Moroshkina, N. (2021). The effect of true and false unreportable hints on anagram problem solving, restructuring, and the Aha!-experience. *Journal of Cognitive Psychology, 33*(6–7), pp. 644–658. https://doi.org/10.1080/20445911.2020.1844722

Arditi, R., & Ginzburg, L. R. (1989). Coupling in predator-prey dynamics: Ratio-dependence. *Journal of Theoretical Biology, 139*(3), pp. 311–326. https://doi.org/10.1016/S0022-5193(89)80211-5

Arditi, R., & Ginzburg, L. (2012). *How species interact: Altering the standard view on trophic ecology*. Oxford University Press.

Bangu, S. (2021). Mathematical explanations of physical phenomena. *Australasian Journal of Philosophy, 99*(4), pp. 669–682. https://doi.org/10.1080/00048402.2020.1822895

[24] The level of idealization of both models can be judged by the de-idealization assumptions each model makes from the most general pattern of predator–prey interaction, e.g., that the predator population harvests resources from the prey population and this harvesting directly leads to an increase in the predators and inversely harms the prey (Holt, 2011).

Batterman, R. W. (2009). Idealization and modeling. *Synthese, 169*(3), pp. 427–446. https://doi.org/10.1007/s11229-008-9436-1

Baumberger, C. (2019). Explicating objectual understanding: Taking degrees seriously. *Journal for General Philosophy of Science, 50*(3), pp. 367–388. https://doi.org/10.1007/s10838-019-09474-6

Baumberger, C., Beisbart, C., & Brun, G. (2017). What is scientific understanding? An overview of recent debates in epistemology and philosophy of science. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 1–34). Routledge-Taylor & Francis.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology, 56*(2), pp. 73–102. https://doi.org/10.1016/j.cogpsych.2007.02.001

Bowden, E. M. (1997). The effect of reportable and unreportable hints on anagram solution and the aha! experience. *Consciousness and Cognition, 6*(4), pp. 545–573. https://doi.org/10.1006/ccog.1997.0325

Bueno, O., & Colyvan, M. (2011). An inferential conception of the application of mathematics. *Noûs, 45*(2), pp. 345–374. https://doi.org/10.1111/j.1468-0068.2010.00772.x

Chakravartty, A. (2017). *Scientific ontology: Integrating naturalized metaphysics and voluntarist epistemology*. Oxford University Press. https://academic.oup.com/book/9711

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*(10), pp. 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*(3), pp. 355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. The University of Chicago Press.

Craver, C. F., & Povich, M. (2017). The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A, 63*, pp. 31–38. https://doi.org/10.1016/j.shpsa.2017.04.005

Currie, A., & Sterelny, K. (2017). In defence of story-telling. *Studies in History and Philosophy of Science Part A, 62*, pp. 14–21. https://doi.org/10.1016/j.shpsa.2017.03.003

De Regt, H. W. (2009). The epistemic value of understanding. *Philosophy of Science, 76*(5), pp. 585–597. https://doi.org/10.1086/605795

De Regt, H. W. (2015). Scientific understanding: Truth or dare? *Synthese, 192*(12), pp. 3781–3797. https://doi.org/10.1007/s11229-014-0538-7

De Regt, H. W. (2017). *Understanding scientific understanding*. Oxford University Press.

De Regt, H. W. (2023). Can scientific understanding be reduced to knowledge? In K. Khalifa, I. Lawler, & E. Shech (Eds.), *Scientific understanding and representation: Modeling in the physical sciences* (pp. 17–32). Routledge. https://doi.org/10.4324/9781003202905-3

De Regt, H. W., & Gijsbers, V. (2017). How false theories can yield genuine understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 50–75). Routledge-Taylor & Francis. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315686110-4/false-theories-yield-genuine-understanding-de-regt-henk-victor-gijsbers

Elgin, C. Z. (2017). *True enough*. MIT Press.

Forbes, C. (2017). A pragmatic, existentialist approach to the scientific realism debate. *Synthese, 194*(9), pp. 3327–3346. https://doi.org/10.1007/s11229-016-1015-2

Frigg, R., & Hartmann, S. (2020). Models in science. In E. N. Zalta (Eds), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2020/entries/models-science/

Ginzburg, L. R., & Jensen, C. X. J. (2008). From controversy to consensus: The indirect interference functional response. *SIL Proceedings, 30*(2), pp. 297–301. https://doi.org/10.1080/03680770.2008.11902131

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), pp. 462–466. https://doi.org/10.1111/1467-9280.02454

Grimm, S. R. (2006). Is understanding a species of knowledge? *The British Journal for the Philosophy of Science, 57*(3), pp. 515–535. https://doi.org/10.1093/bjps/axl015

Grimm, S. R. (2009). Reliability and the sense of understanding. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives* (pp. 83–99). Pittsburgh University Press.

Grimm, S. R. (2010). The goal of explanation. *Studies in History and Philosophy of Science Part A, 41*(4), pp. 337–344. https://doi.org/10.1016/j.shpsa.2010.10.006

Grimmer, H., Laukkonen, R., Tangen, J., & Von Hippel, W. (2022). Eliciting false insights with semantic priming. *Psychonomic Bulletin & Review, 29*(3), pp. 954–970. https://doi.org/10.3758/s13423-021-02049-x

Hardcastle, G. (n.d.). *A problem-solving account of scientific explanation* [Manuscript]. https://philpapers.org/archive/HARAPA-7.pdf

Hattori, M., Sloman, S. A., & Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin & Review, 20*(4), pp. 790–797. https://doi.org/10.3758/s13423-013-0389-0

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. The Free Press.

Hills, A. (2009). Moral testimony and moral epistemology. *Ethics, 120*(1), pp. 94–127. https://doi.org/10.1086/648610

Hitchcock, C. (2003). Of humean bondage. *The British Journal for the Philosophy of Science, 54*(1), pp. 1–25. https://doi.org/10.1093/bjps/54.1.1

Holling, C. S. (1959). The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *The Canadian Entomologist, 91*(5), pp. 293–320. https://doi.org/10.4039/Ent91293-5

Holt, R. D. (2011). Natural enemy-victim interactions: Do we have a unified theory yet? In S. M. Scheiner & R. W. Michael (Eds.), *The theory of ecology* (pp. 125–161). University of Chicago Press.

Jansson, L. (2014). Causal theories of explanation and the challenge of explanatory disagreement. *Philosophy of Science, 81*(3), pp. 332–348. https://doi.org/10.1086/676687

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), pp. 374–419. https://doi.org/10.1016/0010-0285(90)90008-R

Keil, F. C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Sciences, 7*(8), pp. 368–373. https://doi.org/10.1016/S1364-6613(03)00158-X

Kelp, C. (2015). Understanding phenomena. *Synthese, 192*(12), pp. 3799–3816. https://doi.org/10.1007/s11229-014-0616-x

Khalifa, K. (2012). Inaugurating understanding or repackaging explanation? *Philosophy of Science, 79*(1), pp. 15–37. https://doi.org/10.1086/663235

Khalifa, K. (2013). The role of explanation in understanding. *The British Journal for the Philosophy of Science, 64*(1), pp. 161–187. https://doi.org/10.1093/bjps/axr057

Khalifa, K. (2017). *Understanding, explanation, and scientific knowledge*. Cambridge University Press. https://doi.org/10.1017/9781108164276

Khalifa, K. (2023). Should friends and frenemies of understanding be friends? Discussing de Regt. In I. Lawler, K. Khalifa, & E. Shech (Eds.), *Scientific understanding and representation: Modeling in the physical sciences* (pp. 33–50). Routledge.

Khalifa, K., Doble, G., & Millson, J. (2020). Counterfactuals and explanatory pluralism. *The British Journal for the Philosophy of Science, 71*(4), pp. 1439–1460. https://doi.org/10.1093/bjps/axy048

Khalifa, K., Islam, F., Gamboa, J. P., Wilkenfeld, D. A., & Kostić, D. (2022). Integrating philosophy of understanding with the cognitive sciences. *Frontiers in Systems Neuroscience, 16*, p. 764708. https://doi.org/10.3389/fnsys.2022.764708

Khalifa, K., Lawler, I., & Shech, E. (2023). Introduction. In I. Lawler, K. Khalifa, & E. Shech (Eds.), *Scientific understanding and representation: Modeling in the physical sciences* (pp. 1–13). Routledge. https://www.routledge.com/Scientific-Understanding-and-Representation-Modeling-in-the-Physical-Sciences/Lawler-Khalifa-Shech/p/book/9781032054957

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation (Minnesota Studies in the Philosophy of Science, Volume 13)* (pp. 410–505). University of Minnesota Press.

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology, 65*(1), pp. 71–93. https://doi.org/10.1146/annurev-psych-010213-115154

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science, 64*(3), pp. 485–511. https://doi.org/10.1093/bjps/axs012

Lange, M. (2017). *Because without cause: Non-causal explanation in science and mathematics*. Oxford University Press.

Lange, M. (2021). What could mathematics be for it to function in distinctively mathematical scientific explanations? *Studies in History and Philosophy of Science Part A, 87*, pp. 44–53. https://doi.org/10.1016/j.shpsa.2021.02.002
Laukkonen, R., Webb, M. E., Salvi, C., Tangen, J. M., Slagter, H. A., & Schooler, J. (2018). *On selecting ideas: Cognitive and neurocomputational mechanisms of insight* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ez3tn
Laukkonen, R. E., Ingledew, D. J., Grimmer, H. J., Schooler, J. W., & Tangen, J. M. (2021). Getting a grip on insight: real-time and embodied Aha experiences predict correct solutions. *Cognition and Emotion, 35*(5), 918–935. https://doi.org/10.1080/02699931.2021.1908230
Laukkonen, R. E., Kaveladze, B. T., Tangen, J. M., & Schooler, J. W. (2020). The dark side of Eureka: Artifcially induced Aha moments make facts feel true. *Cognition, 196*, p. 104122. https://doi.org/10.1016/j.cognition.2019.104122
Lawler, I., Khalifa, K., & Shech, E. (2023). *Scientific understanding and representation: Modeling in the physical sciences*. Routledge.
Lewis, D. (1973). *Counterfactuals*. Basil Blackwell.
Lipton, P. (2009). Understanding without explanation. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 43–63). University of Pittsburgh Press.
Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology, 31*(2), pp. 168–217. https://doi.org/10.1006/cogp.1996.0016
McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A, 16*(3), pp. 247–273. https://doi.org/10.1016/0039-3681(85)90003-2
Mizrahi, M. (2012). Idealizations and scientific understanding. *Philosophical Studies, 160*(2), pp. 237–252. https://doi.org/10.1007/s11098-011-9716-3
Newman, M. (2014). EMU and inference: What the explanatory model of scientific understanding ignores. *European Journal for Philosophy of Science, 4*(1), pp. 55–74. https://doi.org/10.1007/s13194-013-0075-0
Northcott, R. (2008). Causation and contrast classes. *Philosophical Studies, 139*(1), pp. 111–123. https://doi.org/10.1007/s11098-007-9105-0
Novaes, C. D. (2020). *The dialogical roots of deduction: Historical, cognitive, and philosophical perspectives on reasoning*. Cambridge University Press.
Osuna-Mascaró, A. J., & Auersperg, A. M. I. (2021). Current understanding of the "insight" phenomenon across disciplines. *Frontiers in Psychology, 12*, p. 791398. https://doi.org/10.3389/fpsyg.2021.791398
Park, S. (2017). Understanding without justification and belief? *Principia: an International Journal of Epistemology, 21*(3), 379–389. https://doi.org/10.5007/1808-1711.2017v21n3p379
Park, S. (2020). Scientific understanding, fictional understanding, and scientific progress. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie, 51*(1), pp. 173–184. https://doi.org/10.1007/s10838-019-09480-8
Petkov, S. (2015). Explanatory unification and conceptualization. *Synthese, 192*(11), pp. 3695–3717. https://doi.org/10.1007/s11229-015-0716-2
Petkov, S. (2019). Studying controversies: Unification, contradiction, integration. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie, 50*(1), pp. 103–128.
Petkov, S. (2021). The degrees of understanding and the inferential component of understanding. *Organon F, 28*(4), 746–776. https://doi.org/10.31577/orgf.2021.28401
Pincock, C. (2018). Explanatory relevance and contrastive explanation. *Philosophy of Science, 85*(5), 806–818. https://doi.org/10.1086/699715
Potochnik, A. (2017). Idealization and the aims of science. *University of Chicago Press*. https://doi.org/10.7208/chicago/9780226507194.001.0001
Reiss, J. (2015). *Causation, evidence, and inference*. Routledge.
Rice, C., & Rohwer, Y. (2021). How to reconcile a unified account of explanation with explanatory diversity. *Foundations of Science, 26*(4), 1025–1047. https://doi.org/10.1007/s10699-019-09647-y
Rohwer, Y., & Rice, C. (2016). How are models and explanations related? *Erkenntnis, 81*(5), 1127–1148. https://doi.org/10.1007/s10670-015-9788-0
Rosenzweig, M. L. (1971). Paradox of enrichment: Destabilization of exploitation ecosystems in ecological time. *Science, 171*(3969), 385–387. https://doi.org/10.1126/science.171.3969.385
Salmon, W. C. (1990). Scientific explanation: Causation and unification. *Crítica: Revista Hispanoamericana De Filosofía, 22*(66), 3–23.

Salvi, C., Beeman, M., Bikson, M., McKinley, R., & Grafman, J. (2020). TDCS to the right anterior temporal lobe facilitates insight problem-solving. *Scientific Reports, 10*(1), p. 946. https://doi.org/10.1038/s41598-020-57724-1

Salvi, C., & Bowden, E. (2020). The relation between state and trait risk taking and problem-solving. *Psychological Research Psychologische Forschung, 84*(5), 1235–1248. https://doi.org/10.1007/s00426-019-01152-y

Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., & Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. *Psychonomic Bulletin & Review, 22*(6), 1814–1819. https://doi.org/10.3758/s13423-015-0845-0

Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs, 128*(2), 165–207.

Schunn, C. D., & Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. *Memory & Cognition, 24*(3), 271–284. https://doi.org/10.3758/BF03213292

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135*(1), 94–120. https://doi.org/10.1037/a0014212

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Strevens, M. (2017). How idealizations provide understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 37–49). Routledge-Taylor & Francis.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69*(2), pp. 212–233. https://doi.org/10.1086/341050

van Fraassen, B. C. (1980). *The Scientific image*. Oxford University Press. https://doi.org/10.1093/0198244274.001.0001

Van Steenburgh, J. J., Fleck, J. I., Beeman, M., & Kounios, J. (2012). Insight. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 475–491). Oxford University Press.

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically[1]. *Nature, 118*(2972), pp. 558–560. https://doi.org/10.1038/118558a0

Webb, M. E., Cropper, S. J., & Little, D. R. (2019). "Aha!" is stronger when preceded by a "huh?": Presentation of a solution affects ratings of aha experience conditional on accuracy. *Thinking & Reasoning, 25*(3), pp. 324–364. https://doi.org/10.1080/13546783.2018.1523807

Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy, 104*(12), pp. 639–659. https://doi.org/10.5840/jphil20071041240

Wilkenfeld, D. A. (2017). MUDdy understanding. *Synthese, 194*(4), pp. 1273–1293. https://doi.org/10.1007/s11229-015-0992-x

Woodward, J. (2003). Making things happen: A theory of causal explanation. *Oxford University Press*. https://doi.org/10.1093/0195155270.001.0001

Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science, 2*(3), pp. 375–394. https://doi.org/10.1007/s13194-012-0048-8

Ylikoski, P. (2009). The illusion of depths of understanding in science. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 100–119). University of Pittsburgh Press.