Philosophical Studies
https://doi.org/10.1007/s11098-025-02336-6

A New Account of Pragmatic Understanding, Applied to the Case of AI-Assisted Science

Michael T. Stuart[1]

Accepted: 17 April 2025

Abstract
This paper presents a new account of pragmatic understanding based on the idea that such understanding requires skills rather than abilities. Specifically, one has pragmatic understanding of an affordance space when one has, and is responsible for having, skills that facilitate the achievement of some aims using that affordance space. In science, having skills counts as having pragmatic understanding when the development of those skills is praiseworthy. Skills are different from abilities at least in the sense that they are task-specific, can be learned, and we have some cognitive control over their deployment. This paper considers how the use of AI in science facilitates or frustrates the achievement of this kind of understanding. I argue that we cannot properly ascribe this kind of understanding to any current or near-future algorithm itself. But there *are* ways that we can use AI algorithms to increase pragmatic understanding, namely, when we take advantage of their abilities to increase our own skills (as individuals or communities). This can happen when AI features in human-performed science as either a tool or a collaborator.

Keywords Scientific understanding · Pragmatic understanding · Understanding · Artificial intelligence · Artifactualism · Skills

1 Introduction

Artificial intelligence algorithms are now a permanent part of scientific practice. If philosophers are correct that one main aim of science (if not *the* main aim) is understanding the world, then we need to ask how the use of those algorithms affects the pursuit of understanding. One worry is that the use of AI hinders our pursuit of understanding because AI algorithms are opaque, that is, their inner workings are too complex for humans to grasp and therefore we cannot be sure their outputs are justified (see, e.g., Boge, 2022).

✉ Michael T. Stuart
mike.stuart.post@gmail.com

1 Department of Philosophy, University of York, York YO10 5DD, UK

Published online: 03 July 2025

This paper has three primary goals. The first is to present a more complete picture of the options for discussing AI and scientific understanding. The second is to put forward an account of one under-theorized kind of understanding, namely, *pragmatic understanding*. The final goal is to argue that AI can promote that kind of understanding as a tool or collaborator, but it cannot have such understanding itself.

2 Artificial intelligence and understanding: sketching a more complete picture of the landscape

Accounts of scientific understanding should specify what *kind* of understanding is at issue, what sorts of *objects* understanding is of, what kinds of *agents* understand, and *how* that understanding comes about.

Focusing first on the possible kinds of understanding, we can follow Hannon (2021) in distinguishing at least between *explanatory* understanding (something like grasping a correct explanation), *objectual* understanding (something like grasping a sufficient number of unificatory relations between elements of a domain), and *pragmatic* understanding (something like being able to do something). In terms of objects of understanding, we can follow Shech (2022) in distinguishing at least between understanding a phenomenon, understanding a theory, and understanding why a phenomenon occurs. In terms of the agent who understands, we should at least distinguish between individual humans (who might use AI as a tool, see, e.g., Stuart, 2022), collaborative teams (possibly including AI algorithms as collaborators, see, e.g., Khosrowi et al., 2023), and AI algorithms as agents with their own epistemic goals that they pursue to some extent independently (Barman et al., 2024). Finally, we can also distinguish between the ways that AI might increase understanding, for example, by providing explanations, by representing the target accurately, by simplifying, by producing new data, and by identifying new concepts or conceptual connections.

Combining these variables presents us with many different kinds of question. How might a human-AI *collaboration* gain *objectual* understanding of a scientific *theory* by means of *exploring possibilities*? How might an *AI algorithm* generate *explanatory* understanding of a *phenomenon* by means of *pattern-recognition*? We can also put these questions into particular scientific contexts to produce more specific questions. For example, a medical doctor might use AI as a tool for increasing explanatory understanding of a particular patient's symptoms by predicting a condition based on health indicators (e.g., heartrate, blood pressure, and body temperature), and a climate scientist might use AI as a tool for increasing explanatory understanding of the change in frequency of typhoons based on local sea level and windspeed measurements, and these two uses might require different philosophical accounts, despite the fact that both of them feature an *individual* gaining *explanatory* understanding of a *phenomenon* via *data-driven prediction*.

None of the above question-types, in their general or applied forms, are intrinsically of less philosophical value than the others. This is true despite the fact that most philosophical discussion has concerned the use of AI *as a tool* used by *individuals* to generate *explanatory* understanding of *why a phenomenon occurs* by means of producing an *explanation* (Boge, 2022; Meskhidze, 2023; Räz & Beisbart, 2024; Sullivan, 2022b, 2022c; Tamir & Shech, 2022, 2023). Interestingly, much of the work *outside* philosophy is concerned with how much understanding (of various types, of various objects) can be instantiated or produced within an *AI algorithm* itself (Barman et al., 2024; Krenn et al., 2020, 2022). These two issues represent only a relatively small fraction of the above-described problem-space that could be explored.

Perhaps the reason philosophers have mostly limited themselves to discussing explanatory understanding is because, historically, understanding has been characterized as a side-effect of possessing a good explanation (and as a result, explanation has been the key concept instead of understanding), and because Humphreys' work on opacity and computer simulations (e.g., 2009) provides a nice entry point for discussing how the nature of AI might present a problem for obtaining new, good explanations. It is true that the structure and content of AI algorithms are opaque to the human mind, and yet it seems that they can be used to produce new understanding, and this presents a puzzle which is worth addressing, and can potentially be addressed using existing philosophical frameworks (Stuart & Nersessian, 2019; Sullivan, 2022a, 2022c). For example, Sullivan (2023) uses the epistemology of scientific toy models to explain how AI models (which are highly idealized, just like toy models) can contribute to explanatory understanding.

Explanatory understanding is a fine place to start. But there is room to be more specific about explanatory understanding (explanatory understanding of what, by whom, and how produced?), as well as for considering other kinds of understanding. A restricted focus on explanatory understanding could be justified by appeal to ethical or pragmatic reasons. For example, if curing cancer mainly required new explanatory understanding, then it would make sense to focus on how AI can increase explanatory understanding of how cancer develops in the body. However, it is still unclear how the different kinds of understanding relate to one another. Thus, if explanatory understanding really is the main goal of some scientific field, but increased objectual understanding is required in order to increase explanatory understanding, then the focus could (and perhaps should) shift to objectual understanding. Since we are still at an early stage in the discussion about relations between kinds of scientific understanding, it makes sense to keep an open mind.

This paper will consider the corner of the problem-space that involves *pragmatic* understanding. This is worth doing because (a) much of the way that scientists talk about AI is as a tool that increases their abilities, or what they are able to do, (b) pragmatic understanding is relatively under-theorized, and (c) the results of the discussion will be relevant to whether and how AI algorithms can increase objectual and explanatory understanding.

3 Pragmatic understanding

Pragmatic understanding, also sometimes called practical understanding, seems like a new idea in the context of recent epistemological work on understanding. After all, it was named only recently by Bengson (2017), who claimed that it "is not only absent from most discussions of the nature of understanding; it is also implicitly sidelined or explicitly dismissed" (Bengson, 2020). Bengson is right that previous mentions have mostly been dismissive. Here are two examples Bengson doesn't give. Lipton mentions "procedural understanding" as the sui generis form of understanding provided by abilities, though he explicitly refuses to discuss the sort of understanding involved in such cases (Lipton, 2009; Khalifa, 2013a, 2013b, p. 163). Khalifa likewise admits the existence of such a kind of understanding, only to dismiss it in a footnote: "Note that there is another kind of understanding-how that is of a practical variety, e.g. Jimi understands how to play guitar. This is clearly not explanatory," and *therefore* not to be discussed (Khalifa, 2013b, p. 1164).

Still, of the three kinds of understanding mentioned above, pragmatic understanding might be among the oldest. Zagzebski calls it "understanding-how" and points out that for Plato, "understanding is connected with learning an art or skill, a *technê*. One gains understanding by knowing how to do something well, and this makes one a reliable person to consult in matters pertaining to the skill in question" (Zagzebski, 2008, p. 144). It might also be seen in Aristotle's notion of *technê*, examples of which include skill in shipbuilding, knitting, medicine, gymnastics and rhetoric (Pavese, 2024). It also echoes ideas in Confucian epistemology, such as "knowing to act in the moment" (Hetherington & Lai, 2012; Lai, 2012). But what is it? Hannon's summary is a good place to start:

> Practical understanding…is centrally concerned with skillful action and practical activity. As such, this type of understanding is more closely tied to abilities (i.e., physical dispositions, habits, or bodily activities) than explanations. (Hannon, 2021)

So pragmatic understanding is connected to skill and ability, but what is the nature of that connection? Bengson writes that the *paradigmatic manifestation* of pragmatic understanding is skillful activity, as opposed to "reflexive or instinctive behaviors, mechanical mimicries, and spurts of raw talent or mere knack" (2020). So perhaps skillful action is *evidence of* pragmatic understanding (see also Faye, 2014, p. 34; Stuart, 2016). But evidence is an epistemic notion, not an ontological one. For Bengson, pragmatic understanding is "the systematic, general, practical *grasp* of the skilled agent" (2020). This grasp is a praiseworthy psychological state of the agent. For Bengson, pragmatic understanding is not possessing an ability or skill, but intellectually *grasping a method*, e.g., appreciating how each step of a method leads to the next, appreciating the nuances and leeway in each step of the method, and so on (for development, see Westerblad 2023; Westerblad ms, Kieval and Westerblad ms.). Others postulate a closer connection between pragmatic understanding and ability. Thus, Delarivière and Van Kerkhove (2021) deny that pragmatic understanding should be wholly understood in terms of grasp, and instead identify pragmatic understanding with the possession of sufficient abilities that are contextually appropriate. Currie (2020) and Lenhard (2006, 2009, 2019) likewise identify pragmatic understanding with abilities, e.g., abilities to predict and control. Leonelli characterizes pragmatic understanding as a "cognitive achievement realizable by scientists through their ability to coordinate theoretical and embodied knowledge that apply to a specific phenomenon" (2009).

So, does pragmatic understanding merely *result* in certain abilities (Bengson), or is pragmatic understanding instead *just the having of* certain abilities (Currie, Delarivière and Van Kerkhove, Lenhard), or is having abilities a necessary but insufficient condition for having pragmatic understanding (Leonelli)? Against Bengson, while there is currently no consensus on the nature of grasp, according to at least one recent account, it is just another kind of ability (Strevens, 2024), so his account might reduce to an ability-based account in the end. Second, while people who have pragmatic understanding do often grasp a method in an intellectual way, this seems to be something that comes later in the process. For example, drawing on data from a four-year qualitative study of ecologists, Poliseli shows that epistemic abilities and skills that (at least partially) constitute scientific understanding are often developed *before* the corresponding methods have been made explicit (2020). Indeed, we should expect this to be the case in general, as new methods must be discovered and mastered experimentally before their various steps can be represented and grasped. It seems wrong to exclude such mastery from counting as understanding due to the lack of intellectual grasp of the underlying method. The reason why Bengson claims that pragmatic understanding is to be identified with grasp of a method is because he wants pragmatic understanding to be a distinctively *intellectual* and *epistemic* achievement, as opposed to something like physical ability, which is not necessarily either. However, pragmatic understanding as ability can be intellectual and epistemic, e.g., when the abilities in question are deployed for epistemic ends, or when the abilities are characteristically of an epistemic type (e.g., inferential abilities).

On the other hand, I want to agree with Bengson that pragmatic understanding shouldn't *merely* be identified with the possession of (sets of) abilities either. This is because of the nature of ability. Some philosophers define ability in terms of what would happen in certain contexts (Ginet, 1980). Thus, a concert pianist would play a particular piece of music if they were at a piano and they tried to play it, and this is just what it means to say that they have the ability to play that piece. Others prefer to say that a pianist is "disposed" to play a piece if they tried, and perhaps also that they would have this disposition in the majority of close possible worlds. This ensures that it isn't a fluke that they have this disposition (Fara, 2008; Smith, 2003; Vihvelin, 2013). Still others define ability in terms of an agent doing some particular thing in at least one possible world (Brown, 1988). This makes it "possible" (in the technical sense taken from modal logic) for that agent to do that thing, which might be what we mean when we say that they are "able" to do it. Another account explains abilities in terms of "powers" or "potentialities," which are themselves explained in terms of dispositions (Vetter, 2015; for more recent work on ability and the epistemology of ability, see Vetter and Schoonen, forthcoming).

I follow Bengson in thinking that abilities should not be coextensive with pragmatic understanding because understanding (of any kind) should be seen as a *praiseworthy achievement*. Doing something in a possible world, or having a disposition to do something in this world, are not necessarily achievements. For example, being able to see what's in front of you is an ability, but it shouldn't amount to *understanding* at least because it is not a praiseworthy achievement. Naturally sighted people don't consciously do anything to build that ability. It is also too general: after all, what do people with 20/20 vision understand that people with 20/30 vision don't?

They have more visible details available to them, but having more details available isn't the same thing as having more understanding, just as standing in an archive isn't the same as understanding history.

Still, something about the understanding-as-ability view is on the right track. Consider "expert vision," for example, the ability to see cancer tumours in x-rays, identify forgeries in paintings, or analyze handwriting to determine whether different texts were written by the same person. This is closer to what we're looking for, because it can be learned, it is to some extent under conscious control, and it counts as a praiseworthy cognitive achievement (Stokes, 2021). People who have developed some expertise in seeing certain things understand *those things*, to some extent, and this is true even if they do not grasp the nuances of a particular method for doing what they do.

For reasons like these, Pavese distinguishes between abilities and skills (2024). Skills are a special subset of abilities, namely, they are those abilities that can be learned and mastered, which are characteristically manifested in intentional actions (over which skilled agents can exhibit some cognitive control). These properties of skills help to clarify the difference between what *I* do when looking at an x-ray and what the expert does. I somewhat passively take in information without really *doing* much of anything, while the expert *scans*, *classifies*, and *judges*, which are purpose-driven mental actions that are to some extent under the expert's control. The expert learned how to do this, and they could teach me to do it. But I can only master that skill through intentional practice.[1]

According to Pavese, skills are different from virtues because choosing not to exercise a skill is not a deficiency in that skill (my sister-in-law is still an able chef even if she always orders-in), whereas choosing not to exercise a virtue would be a deficiency in that virtue (if a charitable person always decides not to help others, are they really charitable?). Skills are different from habits and instincts insofar as we cannot control the latter intentionally. Skills also differ from know-how, at least in the sense that you can know how to do something without being skilled in doing that kind of thing. For example, you can take a few swimming lessons and thus know-how to swim, despite not being skilled at it (Pavese, 2024).

I want to add four further points concerning the notion of skill. First, skills are almost never atomistic. Rather, they tend to come in groups. That is, whenever one gains a skill to do one kind of thing, they usually build up several other related skills at the same time. If you learn to play the guitar, for example, you also learn how to read music and identify notes and chords by ear.

Second, the quality of skills can be measured and compared, at least roughly. Thus, a skill is more developed to the extent that it is more robust to changes in the environment, to the extent that it enables (quantitatively) more successful relevant

[1] We do not need to adopt Pavese's view that skills are a kind of ability. Perhaps skills and abilities are simply two different kinds of thing. The important thing for the purposes of this paper is that skills are task-specific, learned and used intentionally, and deployed in a way that can be under cognitive control, which is not necessarily true for all abilities, e.g., the ability to breathe or see.

actions to be performed, or to the extent that those actions which it enables are (qualitatively) more successful, however success is defined.

Third, I want to limit our attention to skills that are at a medium level of specificity. Thus, we might focus on learning to play the guitar, rather than learning to play *that* guitar (too specific). Learning to play the guitar is also learning to play fretted instruments including, to some extent, the bass guitar, banjo, and ukelele. But we will not focus on learning something as general as playing fretted instruments, or learning musicianship (too general). This is important in the present context because we are interested in skills like preparing samples for testing in a wet lab or building computational models or interpreting certain kinds of data, which are skills that are somewhat generalizable, though usually only with caution.[2]

Fourth, in some cases, skills can be *achievements*. For Bradford, an achievement is when an agent intentionally and competently carries out a process that requires a lot of effort to bring about (2015). Consider again the distinction between cases of expert seeing from cases of everyday perception. In the former, the achievement is something specific: facility with the "affordances" (possibilities for action) of a kind of thing, for example, facility with a certain kind of representation (e.g., radiograms) such that something within them can be identified (e.g., cancer tumours). Second, some intentional cognitive effort has gone into developing that skill. And it is that effort which grounds the notion that something has been achieved.

Someone could reply that understanding need not be an achievement, after all, "some instances of understanding are so easy that they require nothing more than simple past experience—for example, understanding a stop sign in the United States" (Zagzebski, 2008, p. 144). There are at least two ways such cases may arise. First, we might have cases in which very little effort was required to produce the relevant skill, but we still think it is permissible to attribute outright understanding because the task that requires that skill is very simple. This would describe cases where an adult first encounters a new kind of traffic sign and learns to identify it without much effort. This might qualify as a deviant kind of non-praiseworthy pragmatic understanding. Such cases are quotidian in science, where new terms are encountered frequently. Of course, in some cases, great effort will be required to build the skills that are required to enable effective use of a new term, concept, or model, and in those cases, genuine pragmatic understanding is the result (for examples, see, e.g., Stuart, 2016, 2018), but in other cases, the scientist merely searches for the meaning of a term online and encounters no difficulty in using the term. In such cases, while there might be a deviant kind of pragmatic understanding there, it is not praiseworthy, and therefore perhaps not the kind of thing that scientists or philosophers will be most interested in accounting for.

A second kind of case is one in which praiseworthy effort *was* necessary to develop the skill, but the agent has since moved into a different context where

[2] Pavese identifies skills with practical knowledge (2024). It would be interesting to consider how practical knowledge differs (if at all) from pragmatic understanding. A similarly interesting comparison would be to Hasok Chang's notion of active knowledge (Chang, 2022). These comparisons are left for future work.

having that skill no longer counts as praiseworthy. Learning to read is very difficult, and comprehending the meaning of a stop sign could qualify as evidence of a praiseworthy achievement for someone who is learning how to read. But when that same skill (reading) is exercised by a working scientist, it is no longer praiseworthy, because the context has changed. What makes this a kind of non-praiseworthy understanding is the agent's being evaluated in the context of professional science, which has higher epistemic standards.

What matters is that in neither case is a mere ability, which need not require any conscious effort to learn and master, sufficient for understanding. Skill is still what counts. If a kind of non-praiseworthy pragmatic understanding is possible, this is either because the tasks faced by the agent are extremely easy, or because the skills required no longer count as praiseworthy due to the higher standards of the relevant context. Cases that meet these conditions in science are not the ones that will be the primary interest of philosophers of science.

With all that in place, here is the account: An agent has pragmatic understanding with respect to some system iff they have the skills to robustly and successfully manipulate that system or its parts to achieve their goal(s), and they are responsible (praiseworthy) for having gained those skills. We can specify this notion of pragmatic understanding in internalist and externalist ways. According to the internalist version, an agent has pragmatic understanding with respect to some system iff they have, *and correctly recognizes that they have*, the skills to robustly and successfully manipulate that system or its parts to achieve their goal(s), *and* they are *responsible* (praiseworthy) for having gained those skills. According to the externalist version, an agent has pragmatic understanding with respect to some system iff they have the skills to robustly and successfully manipulate that system or its parts to achieve their goal(s), *and* they are *responsible* (praiseworthy) for having gained those skills.[3] In both cases, the agent will have more pragmatic understanding to the extent that their skills are better developed.[4]

To illustrate the difference between these two statements of pragmatic understanding, consider a climate scientist who is building an AI model to identify patterns in a particular set of climate data. Let's suppose that the scientist is quite skilled in building AI climate models. According to the internalist, the scientist only counts as having pragmatic understanding if they recognize that they have those skills. According to the externalist, the scientist has pragmatic understanding whether they recognize their own skills or not. This would be relevant in cases where someone very skilled actually believes themselves not to be skilled, which might happen, for example, due to being a member of an underrepresented social group and facing systematic bias, causing feelings of lower self-confidence.

[3] Because skills must be learned through intentional action (practice), responsibility for the achievement of developing a skill is in some sense already part of the definition of what a skill is. However, I have added the responsibility requirement explicitly because it will be important to have it in the forefront of our minds in what follows.

[4] Some might prefer to add an additional requirement: that the goals be epistemic. I want to keep the definition of pragmatic understanding broader than that, but specifying it in this way will not affect any of the arguments to follow.

Another thing worth mentioning about this characterization concerns the inclusion of the agent's goals. Externalists are free to jettison this by characterizing the agent's skills as being relevant for the achievement of some rationally reconstructed goal(s) that we now think have value. Thus, a scientist may possess pragmatic understanding of something even if we now think that they were wrong about what that thing was, or what they should have been trying to do with it. For example, scientists who skillfully manipulated and measured the inputs and products of combustion reactions had pragmatic understanding, even though they may have (mis)characterized their goals as being about, e.g., phlogiston (Chang, 2012).

As usual, some will find the internalist version more appealing, others the externalist version. I won't try to adjudicate between them. Presenting these two different versions is just a way of acknowledging this tension, and giving resources to those who want to think more about it. The difference won't be relevant for what follows.

One last definitional point: pragmatic understanding, like other kinds of understanding, must be *of something*. Skills are task-specific, they come in sets, and they are teachable, but must be practiced for mastery. It may be tempting to say that a skill provides understanding of its characteristic task, but a scientist's skill in producing a certain kind of phenomenon in the lab is not merely an understanding of how to produce that phenomenon. It is also an understanding of what it is possible to do with that phenomenon. This may be made more specific by appeal to the notion of an affordance-space. By "affordance-space" I mean all the affordances (i.e., all the possibilities for action that the agent can recognize) which a thing offers to scientists with various goals. "Affordances" were introduced by Gibson as a technical term to capture the resources provided by an environment for an organism's actions (Gibson, 1979). Since then, there has much discussion concerning the nature of affordances, e.g., do they exist in an environment even when organisms do not? Are they properties or relations? Relations between what? (For reviews, see, e.g., Heras-Escribano, 2019; Chemero, 2003). Luckily, most of the details do not matter for this paper, though they will matter for producing a more complete definition of pragmatic understanding. One pressing detail is that the notion of an affordance must not be defined such that all affordances are always physical, or visually perceivable. This is to make room for things which are sometimes called "cognitive affordances," e.g., things which allow for certain kinds of cognitive actions, like planning, imagining, inferring, etc. (Bruineberg et al., 2019). In sum, we will think of skill in piano-playing as pragmatic understanding of the affordance-space of the piano, in other words, as (a set of) skill(s) concerning the effective use of the possibilities for goal-directed action offered by a piano. Likewise, skill with quantum mechanics is pragmatic understanding of the affordance-space of the concepts and structural (mathematical) features of that theory. Skill with model organisms, like *saccharomyces cerevisiae*, *caenorhabditis elegans* or *drosophila*, is skill concerning the effective use of the possibilities for goal-directed action offered by those organisms.

To summarize so far, we have defined pragmatic understanding as having a skill (or skill-set) which was learned and which manifests itself in intentional action. Because skills are *developed* to some extent *on-purpose*, the agent who develops that skill is responsible (at least in the sense of being potentially praiseworthy) for that skill.

Given this, we can now see how pragmatic understanding relates to the two other kinds of understanding that feature more prominently in the literature, and we can also see why it should count as a kind of understanding. Starting with the latter, notice how pragmatic understanding seems to describe what several philosophers have had in mind when characterizing understanding in general. For Wittgenstein, understanding is the skill to use knowledge (P.I. §§151–155). For Elgin, understanding involves "a capacity to operate successfully within the constraints the discipline dictates or to challenge those constraints effectively. And it involves an ability to profit from cognitive labors, to draw out the implications of findings, to integrate them into theory, to utilize them in practice" (1993, pp. 14–15). It is not "a matter of believing…It involves knowing how to wield one's commitments to further one's epistemic ends. It involves being able to draw inferences, raise questions, frame potentially fruitful inquiries, and so forth" (2017). For Potochnik, "Genuine understanding…requires successful mastery, in some sense, of the target of understanding" (2017, p. 94). For Le Bihan, understanding is "a cognitive success" that "manifest[s] itself through some abilities, including abilities to infer, generalize, transfer, and answer [what if things were different]-questions" (2017). What unites these conceptions is that understanding somehow involves cognitive competence, skills or abilities that allow one to "wield" things to "further one's epistemic ends" (Elgin, 2017). What I have tried to offer is a way of spelling out this intuition in the language of skills, affordances, and praiseworthiness. The fact that these philosophers characterize understanding in this way is, prima facie, a reason to think that pragmatic understanding is indeed a kind of understanding.

We can now (very briefly) turn to the relationship between pragmatic understanding and the other two kinds of understanding. Those interested in explanatory understanding (including Pritchard, 2010; Hempel, 1965; Kitcher, 1989; Grimm, 2006; Khalifa, 2012, 2017; Strevens, 2013; Hills, 2016; de Regt, 2017), portray understanding as something like grasping a correct explanation. For example, Strevens claims that understanding is grasping an explanation which lays out the causal history of a phenomenon and "strips away" anything that is not a "difference maker" (2013). For Khalifa, understanding is grasping the "explanatory nexus" in a way that resembles scientific knowledge (2017). For de Regt, someone understands something if and only if they possess an explanation of that thing, which is based on an intelligible theory (which is "one that has a cluster of qualities that facilitate its use for a given scientist") and conforms to the basic epistemic values of empirical adequacy and internal consistency (2017, p. 92). Most of these accounts invoke a notion of "grasp," which is important, because merely *hearing* the words of a correct explanation doesn't seem sufficient to produce understanding. Rather, the agent must also "attend to" the explanation, and see how the explanans explains the explanandum, in a way that enables the agent to interact successfully with that phenomenon, e.g., by being able to make new inferences and answer questions about closed related phenomena (Hills, 2016). As we noted above, this grasping might be best characterized as the exercise of, or development of, a skill. For example, Grimm argues that grasp is an ability related to manipulating counterfactuals (2006), and Strevens claims that grasp is a recognitional ability, specifically, the ability to grasp a property, which is "to have a great proficiency in tasks related to that property" (2024).[5] As I argued above, at least in the scientific context, it is better to speak in terms of skills here, rather than mere abilities, since, e.g., no one is born with conceptual abilities relating to modern scientific concepts like superposition, enthalpy, and the Krebs cycle. The upshot is that if we think of grasp in this way, explanatory understanding will require pragmatic understanding.

In a bit more detail, we might consider Khalifa's account of explanatory understanding, which allows that "although minimal understanding clearly involves no special abilities, it does involve some abilities" (2017, pp. 59–60). One relevant ability concerns the possession of concepts, which is necessary for grasping explanations. He writes, "to possess a concept is to be able to use it correctly…one person has greater mastery of a concept if the former can use that concept in more correct ways than the latter; e.g., mastering the different roles it can play in different inferences" (2017, p. 59). On this account it is clear that conceptual mastery is a matter of degree, and we might think that in some cases, including scientific cases, high degrees of conceptual mastery should be considered not just as abilities, but also as praiseworthy skills. And such skills are required, on Khalifa's account, for scientific explanatory understanding.

Or consider de Regt's account, on which possessing an explanation is (necessary and) sufficient for understanding, as long as that explanation is based on an intelligible theory. What it means for a theory to be intelligible is defined in terms of the abilities of the scientist who wants to use a theory. These abilities are central for de Regt's view, as he writes, "If S wants to explain a phenomenon on the basis of [a theory], she needs appropriate skills to use [that theory] for model construction," and without these skills, the theory will not be intelligible, and thus there can be no explanatory understanding (2017). The relevant skills are things like being able to construct models from a theory which then serve as good explanations. On the definitions given above, such abilities should be counted as praiseworthy skills. Thus, for de Regt as well, it seems that at least in some cases, pragmatic understanding will be necessary for explanatory understanding.

Those interested in objectual understanding (including Baumberger, 2011; Baumberger and Brun, 2016; Dellsén, 2020; Kvanvig, 2003; Elgin, 1993, 2017; Wilkenfeld, 2017, 2019; Kelp, 2015), claim that understanding a phenomenon or subject is grasping the dependency relations that unite some relevant domain. Those dependency relations might be causal, mathematical, logical, semantic, or explanatory. For example, you might understand climate change by grasping how certain factors (like the level of CO_2 in the atmosphere) affect global average temperature (Baumberger, 2019), or you might understand automotive repair by grasping how interventions on various car parts will affect the car's functions.

Those who focus on objectual understanding are usually even happier than the explanationists to admit the importance of skills. For example, Elgin is explicit that objectual understanding is in some sense constituted by having certain skills (2017). Another account of objectual understanding is Wilkenfeld's. His earlier ("manipulationist") account of objectual understanding claimed that an agent possesses

[5] For further argument along these lines, see Carter et al. (2021).

understanding when they have mental representations that they are able to "modify in small ways" to produce new representations that enable the understander to draw efficacious inferences about some object (2017). His more recent ("understanding as compression") account likewise makes inferential and representational skills central for understanding, alongside having an appropriate mental representation. This time, the agent must be able to unpack compressed mental representations in inferentially useful ways (2019). As above, this skill might be produced without much intentional effort or deployed without cognitive control. In such cases, objectual understanding will only require non-praiseworthy pragmatic understanding. But in some cases, especially those in science, the skills will be praiseworthy. For example, unpacking Einstein's field equations to predict the existence of black holes required many praiseworthy skills. In general, some pragmatic understanding will be necessary to have objectual understanding of general relativity, as well as for relativistic phenomena. And much else.[6]

In sum, pragmatic understanding has been defined as having the skills required to robustly and successfully manipulate some system or its parts to achieve some goal, while also being responsible for having those skills. I argued that this should count as a kind of understanding, and also that in some cases, it will be the kind of understanding that is necessary for having explanatory or objectual understanding.

We now turn to the question of whether and how AI can increase pragmatic understanding. To do so, this question must be specified, as noted above. Accordingly, the next section considers three cases. In 3.1 we consider whether algorithms *themselves* can possess pragmatic understanding. In 3.2 and 3.3 we consider whether and how AI algorithms can assist as collaborators, or as tools.

4 AI and pragmatic understanding

4.1 AI as agent

On the pragmatic understanding-as-ability view that we rejected above, AI algorithms could possess understanding, since abilities might correctly be attributed to them. This position seems to be endorsed by Barman et al. (2024), who claim that "scientific understanding is an ability and should therefore be measured in terms of behavioral competence (i.e., actions)." They define a benchmark which is meant to be general enough to evaluate the level of understanding of either a human or an AI algorithm, as follows:

> The degree to which agent A scientifically understands phenomenon P can be determined by assessing the extent to which (i) A has a sufficiently complete representation of P; (ii) A can generate internally consistent and empirically

[6] There might be philosophers who would deny that skills are required for or partially constitutive of objectual understanding. But as long as their account requires grasp, the above argument connecting grasp and skill can be made again here. See, e.g., Dellsén (2020), who portrays objectual understanding in terms of grasping a dependency model.

adequate explanations of P; (iii) A can establish a broad range of relevant, correct counterfactual inferences regarding P.

If we are going to set a general benchmark for pragmatic understanding, it is natural to do this in terms of actions that the agent can perform, since abilities and skills manifest themselves through actions. This kind of thinking motivated the Turing test and continues to appear in benchmarks set by AI companies like DeepMind and OpenAI.

But there are several reasons to worry about this particular benchmark. One is that it requires AI algorithms to draw on representations of target phenomenon. Several philosophers have denied that the algorithms used in science do in fact employ such representations (Boge, 2022, Kieval, forthcoming). Second, it is not clear whether behavioural tests can tell us about the existence, content, or quality of an agent's representation, when that agent is an AI (see Delarivière & Van Kerkhove, 2021). Third, concerning condition (ii), it might be the case that what should be required isn't merely an ability (which most advanced LLMs have, e.g., to produce expressions in native-sounding English), but a *skill*, for example, the task-specific ability to craft good scientific explanations of a particular kind, developed through effortful practice. Regarding condition (iii), again, this might be an ability which most AI can satisfy, depending on how we measure breadth. But this is a mere ability, not a praiseworthy skill. Pragmatic understanding as characterized in Sect. 2 requires intentional, praiseworthy behaviour. If AI algorithms are not capable of the intentional action required for responsibility, then they cannot possess pragmatic understanding. And if that's correct, there will be many cases, especially in science, where those algorithms could not possess either explanatory or objectual understanding either, because possessing either of those would require possessing some pragmatic understanding. This is a strong claim, so let's consider it in a bit more detail.

Whether AI algorithms are capable of intentional action or responsibility is a contentious issue in the philosophy of AI. Starting with responsibility, we can identify two general conditions for an agent to be responsible. The first is an epistemic condition, according to which an agent must be "aware" of the consequences of their actions. The second is a control condition, according to which an agent must have "control" over their actions (see, e.g., Mele, 2010; Rudy-Hiller, 2022). There are levels of responsibility, and an agent is *more* responsible to the extent that they are more aware of the consequences of their actions, and/or have more control over their actions. Thus, someone who commits a crime is responsible to the extent that they were aware of what they were doing and could have done otherwise. Someone who is negligent is less responsible than someone who had intentionally committed the same crime because they were not aware of what they were doing (or the consequences of their actions) but they *should have been*. The same goes for praise: a scientist who is less aware of the consequences of their actions and had less control over what they were doing is less responsible for any resulting scientific progress than someone who was fully aware of what they were doing and could have done otherwise.

Is an algorithm "aware" of the consequences of its "actions"? Does it have "control" over its actions? The folk often think so, but philosophers tend to think that this is a mistake which can be traced to our human tendency to anthropomorphize (Kneer & Stuart, 2021; Shevlin & Halina, 2019; Stuart & Kneer, 2021). Algorithms cannot be said to be aware of the consequences of their actions because algorithms are rules for logical operations running more or less determinately. Rules are not agents. In addition, even if they were, their entire "world" is made up of symbols that are pure syntax, at least in the sense that they are not connected to real-world objects. When AlphaFold outputs a hypothesis for the three-dimensional structure of a protein given information about that protein's amino acids, it is not aware of what a protein is, or what an amino acid is, or what protein structure is, or what space or spatial dimensions are. This goes some way towards preventing us from saying that it is aware of what it is doing.

The reason algorithms are not typically said to be in control of their actions is because to have control over one's actions, minimally, one must be able to *act*. Again, algorithms are sets of rules that define processes executable in a computer: they are not agents capable of performing intentional actions, which are actions that spring from (or are plausibly reconstructable as springing from) purposes, reasons, desires, or intentions. This is because purposes, reasons, desires, and intensions are complex mental states with content about the external world, which AI doesn't have, at least until the symbol grounding problem is solved. But again, this is controversial: some might claim that the symbol grounding has been (or will soon be) solved, or that responsibility doesn't require it to be solved.

This is not the place to argue that AI is incapable of satisfying the conditions for responsibility (and thus for possessing pragmatic understanding). Instead, I will simply go along with the majority view in the philosophy of AI, which is that current and near-future AI is not the kind of thing which can bear responsibility for its "actions" (Burton et al., 2020; Hakli & Mäkelä, 2019; Leveringhaus, 2016, 2018; Nyholm, 2018; Sparrow, 2007). If this is fair, then at least for now, AI algorithms, considered on their own, cannot possess pragmatic understanding, and thus AI algorithms also will not be able to possess explanatory and objectual understanding (at least insofar as those require skills like conceptual mastery/grasping).

This is surprising, because we might have thought that AI algorithms could possess (in some sense) explanatory or objectual understanding because AI can "possess" (in some sense) explanations or dependency models. And we also might have thought that AI algorithms could possess pragmatic understanding, as it seems that they might qualify as having abilities or as having grasping a method. But, as we've just seen, that is not always going to be enough. Abilities might be programmed, just as they sometimes are by evolution. But skills cannot be programmed, and it is skills that are required for scientific understanding.[7]

[7] A potentially interesting direction for future research would be to consider ways in which we might expand our concept of (pragmatic) understanding to include non-human intelligent systems. One way would be to drop the responsibility requirement, e.g., by defining understanders as any agents that have certain abilities, including inferential abilities. The problem here is that many systems already exist

4.2 AI as collaborator

If AI algorithms are not (yet) fully-fledged agents capable of being responsible for their own capacities then it does not make sense to think of them as collaborators in the usual way. However, as some philosophers have recently pointed out, there are different ways of thinking about what collaboration is. And this is important, given that scientists and artists do sometimes say that interacting with AI is collaborative, even when they appreciate that AI is not agentive (Hertzmann, 2020), and it would be good to be able to make sense of this.

For inspiration, we may look briefly at the case of *artists* who use AI. There are many artists claiming that AI is not a mere tool but also a collaborator (Chung, 2019; Colton, 2012; McCormack et al., 2020). They also claim that AI is a *better* collaborator when it's not very autonomous, that is, when it does pretty much exactly what you ask it to: it produces only "controlled uncertainty" (Miller et al., 2020). What is going on here?

Anscomb (2024) distinguishes between *collective authorship* (which requires mutual responsiveness, a meeting of minds), *co-creatorship* (which allows for separate but still intellectually responsive and creative contributions) and *co-production* (in which someone takes the lead and gives limited freedom to others to carry out subtasks). Looking carefully at each possibility, Anscomb concludes that AI does not meet the conditions required to satisfy any of these. So why do artists speak this way? Perhaps they are reporting feelings that are relevant to them and their peers/audience about the experience of creating art with AI assistance: it *feels* like collaboration, and they think this is worth thinking about. Or perhaps they want to make clear feelings of diminished responsibility: they do not feel that all the ideas represented in their work were entirely produced by them.

This last possibility helps us to pivot from the question of collaboration to one about credit assignment. Anscomb notes that credit assignment concerns who has freedom to make decisions about various aspects of the work (Anscomb, 2021). For example, a head or lead artist can choose to change the main idea motivating the artwork, while the assistants and technicians can only choose to change some very small details. This reflects an asymmetry of power, and also of skill: the head artist typically *could* do what the technicians are doing, but not vice-versa. And the artist also has additional skills, for example, those related to setting projects and (re)interpreting artistic constraints of various kinds (stylistic, material, spatial, etc.), which

Footnote 7 (continued)
whose abilities far outstrip our own, including calculators, telescopes and particle colliders, and it does not seem right to attribute to them any epistemic achievement. On the other hand, we could maintain a need for responsibility, but we might think about AI systems that could meet that requirement by "building up" the requisite responsibility by combining many artificial agents, each of which possesses some small amount of responsibility. But it's not clear that responsibility does aggregate: e.g., a large group of ants, dogs, or babies might together manage to cause some harm, but would nevertheless not be tried in court as a human adult, no matter how many millions of agents were involved. Another strategy would be to focus on the components of responsibility, e.g., control and awareness, and see if AI-friendly analogues of these could be created.

the assistants and technicians do not have. And this is why the bulk of the credit for the project goes to the head artist. In a similar way, an AI algorithm can do some of the things an artist would do, like controlling a robot arm that sketches with a pencil on paper, composing a variation on a musical theme, etc. As this must feel quite similar to the case where certain aspects of the artistic process are outsourced to human assistants and technicians, it explains why some artists might be tempted to use the language of collaboration, even if that might be technically incorrect.

Still thinking about credit assignment, we can think of collaboration in a more functional way: collaborators might be *anything* whose contributions to a project have some minimal value of certain kinds. Thus, Khosrowi et al. (2023, 2024) argue that an agent's share of credit should be determined by how *relevant* their contribution is (i.e., how much of a difference their contribution makes to the final output), how *non-redundant* it is, how much *control* the agent has over the process (i.e., how much the agent could cause the creation process to go in different ways, whether they actually exercise that power or not), how *original* the contribution is, how much *time* and *effort* the agent puts in, how much *leadership* the agent exhibits, how *independent* the agent is in their work, and how *directly* their work contributes to the content of the final output.

AI algorithms are interesting because they may score high on some of these criteria (e.g., their contributions can be relevant, non-redundant, and direct), and low on others (e.g., leadership and effort). None of these conditions are meant to be necessary or sufficient for counting as a member of a collaboration, and the set is not meant to be exhaustive. And we should expect these conditions to come together in ways that are standard for a practice. For example, agents who play leadership roles *in art* often deserve the most credit for their relevant, direct, original, long-term and effortful independent impact, while agents who play leadership roles *in science* often get the most credit for being relevant and non-redundant, despite being less directly involved, expending less time and effort, and not always being the main source of originality.

Since AI can arguably rate highly on several of these criteria, we can follow Khosrowi, Finn and Clark in accepting AI as a collaborator at least in this sense, even if we reject the idea that AI algorithms could be collaborators in the sense of being autonomous agents responsible for their creative work.

Another (compatible) way to think about AI algorithms as collaborators is to use an extended or distributed cognition framework. For example, Nersessian has developed what she calls the "d-cog" framework to analyze cognitive-cultural processes and specifically the collaborative work that takes place in scientific laboratories (Nersessian, 2022, pp. 8ff). Here, we allow for cognitive acts (like representing, remembering, calculating, and imagining) to be distributed over agents and artifacts, and we focus on the way that both can work together in problem-solving, where the problems and the context are constantly changing. We can identify two different ideas here: (1) The lab is "made of" individuals and objects, but since it is impossible to disentangle who is responsible for each action, we should allow the responsibility for progress to bleed across all the individuals and their tools, or (2) The laboratory itself is a single entity over and above its members and artifacts. Thus the lab *itself* discovers, manipulates, develops skills, and understands.

Looking at the first idea, AI algorithms currently possess remarkable abilities that far surpass those of any human. To repeat, they are “mere” abilities, like the ability of an eye to see or a slime mould to find efficient pathways to resources: they do not properly deserve credit, as they aren’t grounded in anything intentional. On their own, these abilities cannot provide any skills to the group, because no level or number of abilities can stack to provide the responsibility required to possess a skill and thus to possess pragmatic understanding. But human individuals in the group can have (or can justifiably take) responsibility for the outputs of the lab’s members and instruments. So, pragmatic understanding can be the result of a combination of the mere abilities of AI algorithms and the responsibility and skills of the human lab members.

Thus, according to the definition of pragmatic understanding given above (glossing over the distinction between the internalist and externalist readings), a laboratory will have pragmatic understanding with respect to some system iff at least some of the members of the lab have the skills to robustly and successfully manipulate that system or its parts to achieve some goal(s), and at least some of the members are responsible for having those skills. For example, any lab that uses AlphaFold has some pragmatic understanding of protein folding since there are abilities and skills distributed among the scientists and their instruments, and to that fuzzy assemblage, we can attribute pragmatic understanding.

The second idea is that the lab itself metaphysically emerges as a new thing, separate from its members. Its cognitive states or actions or skills exist over and above those of its members. In any case, it is the lab itself which is responsible for building the relevant skills, e.g., in the AlphaFold case, to robustly and successfully manipulate representations of proteins to predict protein folding structure. How? Groups can have (or take) responsibility, at least in the legal context. For example, companies can be convicted of crimes. So we might think there could be groups (that are "made of" people but also instruments) which are praiseworthy for building skills that none of the members possess. This would require committing to a controversial metaphysical position according to which all the relevant states and processes relevant for attributing understanding are all attributable only to the group itself.

In sum, there are at least two ways in which it might be possible to think of AI algorithms as collaborators: in a functional way (such that their contributions have value that is relevant, non-redundant, and direct) and a distributed way (such that they constitute parts of a whole to which we attribute certain epistemic properties). In either case, it should be clear how the pragmatic understanding of a group can be increased via the inclusion of programmed abilities. This can then be helpful for explaining how objectual and explanatory understanding can be increased in groups, insofar as those kinds of understanding depend on pragmatic understanding.

4.3 AI as tool

Scholars working on the use of AI in art are split between characterizing AI as a tool and a collaborator, with some proposing that AI might require rethinking this distinction entirely, so that AI falls somewhere in between (Khosrowi et al., 2023). However, it is also possible that some AI algorithms function as a tool, and others function as a collaborator, or even that the same algorithm might in some contexts function as a tool, and in others function as a collaborator. Whether a given AI algorithm "really is" a tool or a collaborator can be left to one side for present purposes, because the question I want to ask is about which epistemological framework is best for making sense of particular uses of AI in science to increase understanding.

So, are there AI algorithms that are sometimes best characterized as tools? This depends on what a tool is.

The idea that tools are necessary for science goes back a very long time, with Francis Bacon stating that.

> Neither the bare hand nor the unaided intellect has much power; the work is done by tools and assistance, and the intellect needs them as much as the hand. As the hand's tools either prompt or guide its motions, so the mind's tools either prompt or warn the intellect. (Bacon, 1620/2000, trans. Jardine and Silverthorne, p. 33)

Bacon distinguished between tools of the mind and tools of the hand, as a way of distinguishing between tools that extend the power of the human mind and body. Following Stuart (2022), we can distinguish between other types of tools, including tools of the senses (which produce information grounded in measurements), and tools of the voice (which assist in presenting and disseminating findings). Tools of the senses may be subdivided into tools of various senses (like tools of the eye, the skin, etc.), just as tools of the mind can be subdivided into tools that assist with particular processes like calculation, inference (e.g., parameter estimation, approximating solutions, curve-fitting, etc.), imagination, memory, and so on.

AI algorithms can be characterized as various kinds of tool. As a tool of the senses, AI can produce higher-resolution images of black holes (Medeiros et al., 2023) or dark matter simulations (Li et al., 2021). As a tool of the voice, AI can assist in public relations writing and marketing, or even (controversially) as part of article writing. It can be a tool of the hand, as when AI is connected to robotics and used to power experimental systems like Adam, Eve and Genesis, which are robotic-AI systems that design and perform experiments semi-autonomously (King et al., 2023).

One reason to treat AI algorithms as tools rather than as something else concerns the fact that AI algorithms, at least the ones used in science, might not *represent* anything. "There just doesn't seem to be any plausible way to use them as scientific representations in most cases" (Kieval forthcoming). If this is correct, we have reason to follow Kieval in adopting an *artifactualist* view of AI algorithms as tools. The idea would be that AI algorithms should not be judged in terms of their accuracy as representations (contra Sullivan, 2023), but in terms of their fitness for purpose.

A hammer is not typically a good representation of anything, but it is very good for hammering. Taking up this perspective, the connection to pragmatic understanding might go something like this. In the same way that a hammer enables an agent to increase the power and focus the power of their hammering actions, AI algorithms enable agents to increase the power and focus of their cognitive or physical abilities. Insofar as a human agent is responsible for learning to build or use such tools, the agent counts as developing skills with those tools, and thus as having some pragmatic understanding of those tools. The agent may also gain skills, indirectly, to manipulate a target system using that tool, in which case, pragmatic understanding of that target is also achieved. Learning to build and use AI models of natural systems can therefore be a way to increase pragmatic understanding, not just because we have grasped new methods, but because we end up with skills concerning the affordance space of tools and the systems on which those tools are used, which enable the achievement of certain goals that we (or our community) regard as important.[8]

But even if AI algorithms do represent their targets, they can still be thought of as tools, since acts of representation in science are typically done for the sake of surrogative reasoning. We build representations because they are cognitively useful. For example, we use representations in science to draw further inferences about systems-as-represented, which might tell us something about theory or reality. Thought of in this way, we can distinguish between representations and representational tools. If AI algorithms are or contain representations, those representations can be evaluated in terms of their epistemic value, e.g., as being accurate reflections of their targets or having good downstream epistemic consequences. But insofar as these representations are being used to do something particular, for example, to explore theory-space or to inspire a hypothesis about a target system, those representations can also be evaluated in terms of their fitness for such purposes. Stuart argues that a consequentialist epistemological framework might be best for making sense of such evaluations (2022): A tool is good when using it has good consequences. One of those good consequences might be increased pragmatic understanding, which can be achieved by developing skills to use tools that productively extend our abilities.

In sum, AI algorithms, whether they are representational or not, can be thought of as tools that extend and empower human abilities. Scientists learning to use those tools develop praiseworthy new skills. When this happens, the result is an increase in pragmatic understanding, primarily of the affordance space of the tool, but surrogatively or derivatively of whatever that tool is deployed to help with, including real-world or fictional-theoretical target systems. This kind of pragmatic understanding can then be the ground for objectual or explanatory understanding of those same systems.

[8] Kieval (forthcoming) is also concerned to show that AI algorithms can be understood from the perspective of skills or skill-building, though he takes his account of skill from Fridland (2021), and doesn't make the connection to a particular account of pragmatic understanding explicit. In other work, Kieval and Westerblad (ms) connect AI-related skills to an account of pragmatic understanding characterized as a set of methodological principles. I worry that identifying pragmatic understanding with methodological principles will allow for the content of pragmatic understanding to be propositional, which will then require a "third thing" that connects those principles to coherent efficient action (e.g., grasp of such principles, or further skills), and also that this distances us from the intuitions mentioned above supporting the idea that pragmatic understanding *is* praiseworthy skill possession. In other work, however, Westerblad (ms) allows for an identification of pragmatic understanding with skill, and claims only that methods are vehicles for pragmatic understanding.

5 Conclusion

The last three decades have seen a strong push to change the way we think about the epistemic aims of science, with some arguing that understanding (rather than knowledge) should be considered *the* goal of science. The fate of this claim will depend on whether we can carve out a meaningful notion of understanding that is different from other central epistemic goods like truth, knowledge, accurate prediction, problem solving power, fruitfulness, and so on, and whether this notion allows us to better describe and evaluate the practice of science. This paper has introduced an account of one kind of understanding that is currently under-theorized, namely pragmatic understanding, in terms of the praiseworthy development of skills.

The last two decades have *also* seen a strong push to (re)focus philosophical attention on the use of digital methods in science, and more recently, on the use of AI algorithms. It is natural, therefore, to explore how the more specific question of how AI contributes to scientific understanding. So far, much of that attention has concerned whether AI algorithms stand in the way of explanatory understanding, given their opaque nature.

This paper has attempted to shed new light on the issue of whether AI algorithms can promote understanding. It did this by expanding the problem-space to include different kinds of understanding, as well as different kinds of agents and objects of understanding. Focusing on one under-studied kind of understanding, namely pragmatic understanding, we considered whether AI algorithms could possess pragmatic understanding on their own (no), whether AI algorithms could increase the understanding of a mixed human-AI group (yes, as collaborators in a merely functional sense or by lending their abilities to a group), or whether they can be thought of as tools which increase the understanding of human agents that use them (yes). Relations were considered to explanatory and objectual understanding, and it was argued that in some cases pragmatic understanding will be necessary for explanatory and objectual understanding. Since AI algorithms themselves cannot possess pragmatic understanding, we should not expect AI algorithms to possess those other kinds of understanding either, though we can expect AI algorithms to make those kinds of understanding possible when contributing to collaborations or serving as tools.

Acknowledgement I have been discussing some of the ideas in this paper for more than a decade now, so I am bound to forget to thank some of the people who have helped along the way. Thanks to Jan Michel for the invitation to submit to this special issue, and to the audience at a workshop he organized called "Artificial Researchers and Scientific Discoveries" at the University of Düsseldorf. Thanks also to the organizers and audience at the Notre Dame Annual HPS Conference on Science and the Imagination (2025), the "Lingnan-Cambridge Workshop on AI in Science" at Cambridge, as well as audiences at: Exeter (2025); Dubrovnik (2025); Stockholm (2024); the British Society for Philosophy of Science (2021 & 2022); the 3rd meeting of the Scientific Understanding and Representation workshop (2021) in Radboud; the Understanding Progress in Science & Beyond at the University of Iceland (2020); Linköping (2020); and the Capstone Conference for the Varieties of the Understanding Project at Fordham (2016). Thanks especially to Christoph Baumberger, Claus Beisbart, Alexander Bird, James Robert Brown, Marco Buzzoni, Elisabeth Camp, Hasok Chang, Adrian Currie, André Curtis-Trudel, Finnur Dellsén, Eamon Duede, Kate Elgin, Roman Frigg, Seth Goldwasser, Till Grüne-Yanoff, Marta Halina, Michael Hannon, Alison Hills, Jonathon Hricko, Vincent Hsu, Milena Ivanova, Mikeal Janvid, Max Jones, Atoosa Kasirzadeh, Kareem Khalifa, Donal Khosrowi, Johanes Lenhard, Sabina Leonelli, Ying-Tung Lin, Chris McCarroll, Nancy Nersessian, James Nguyen, Rune Nyrup, Henk de Regt, Niall Roe, Joe Roussos, Tom

Schoonen, Yafeng Shan, Elay Shech, Alison Springle, Noah Stemeroff, Hugu Sutej, Adam Toon, Philippe Verreault-Julien, Marcel Weber, Oscar Westerblad, Daniel Wilkenfeld, Harald Wiltsche, and Karen Yan. Thanks also to the students in my Philosophy of AI masters programme (2024-2025), and to an anonymous referee at Philosophical Studies.

References

Anscomb, C. (2021). Visibility, creativity, and collective working practices in art and science. *European Journal for Philosophy of Science, 11*(1), 1–23. https://doi.org/10.1007/s13194-020-00310-z

Anscomb, C. (2024). AI: Artistic collaborator? *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02083-y

Bacon, F. (1620) 2000. *The New Organon*. Translated by Lisa Jardine and Michael Silverthorne. Cambridge: Cambridge University Press. https://www.cambridge.org/gb/academic/subjects/philosophy/philosophy-texts/francis-bacon-new-organon; https://www.cambridge.org/gb/academic/subjects/philosophy/philosophy-texts.

Barman, K. G., Caron, S., Claassen, T., & de Regt, H. (2024). Towards a benchmark for scientific understanding in humans and machines. *Minds and Machines, 34*(1), 6. https://doi.org/10.1007/s11023-024-09657-1

Baumberger, Christoph. 2011. "Understanding and its Relation to Knowledge," in *Epistemology: Contexts, Values, Disagreement* (Proceedings of the 34th International Ludwig Wittgenstein Symposium), ed. Christoph Jäger and Winfried Loeffler (Germany: Ontos Verlag), pp. 16-18.

Baumberger, Christoph and Georg Brun. 2017. "Dimensions of Objectual Understanding," in *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, ed. Stephen Grimm, Christoph Baumberger, and Sabine Ammon (New York, NY: Routledge), pp. 165-189

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

Bengson, J. (2017). The unity of understanding. In S. R. Grimm (Ed.), *Making sense of the world: New essays on the philosophy of understanding.* Oxford University Press.

Bengson, J. (2020). Practical understanding: Skill as grasp of method. *Concepts in thought, action, and emotion.* Routledge.

Boge, F. J. (2022). Two dimensions of opacity and the deep learning predicament. *Minds and Machines, 32*(1), 43–75. https://doi.org/10.1007/s11023-021-09569-4

Bradford, G. (2015). *Achievement*. Oxford University Press. https://academic-oup-com.libproxy.york.ac.uk/book/9782

Brown, M. A. (1988). On the logic of ability. *Journal of Philosophical Logic, 17*(1), 1–26.

Bruineberg, J., Chemero, A., & Rietveld, E. (2019). General ecological information supports engagement with affordances for 'higher' cognition. *Synthese, 196*(12), 5231–5251. https://doi.org/10.1007/s11229-018-1716-9

Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P., & Porter, Z. (2020). Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence, 279*(Feb), 103201. https://doi.org/10.1016/j.artint.2019.103201

Carter, J. A., Gordon, E. C., & Grodniewicz, J. P. (2021). Understanding a Communicated Thought. *Synthese, 198*(12), 12137–12151. https://doi.org/10.1007/s11229-020-02854-2

Chang, H. (2012). *Is water H2O?* Vol. 293. Boston Studies in the Philosophy of Science. Springer Netherlands. https://doi.org/10.1007/978-94-007-3932-1.

Chang, H. (2022). *Realism for realistic people: A new pragmatist philosophy of science*. Cambridge University Press.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181–195. https://doi.org/10.1207/S15326969ECO1502_5

Chung, Sougwen. 2019. *Artefacts*. https://sougwen.com/project/artefact1.

Colton, S. (2012). The painting fool: Stories from building an automated painter". *Computers and creativity* (pp. 3–38). Springer.

Currie, A. (2020). Bottled understanding: The role of lab work in ecology. *The British Journal for the Philosophy of Science*, *71*(3), 905–932. https://doi.org/10.1093/bjps/axy047

de Regt, H. (2017). *Understanding Scientific Understanding*. Oxford University Press.

Delarivière, S., & Van Kerkhove, B. (2021). The mark of understanding: In defense of an ability account. *Axiomathes, 31*(5), 619–648. https://doi.org/10.1007/s10516-020-09529-0

Dellsén, F. (2020). Beyond explanation: Understanding as dependency modelling. *The British Journal for the Philosophy of Science, 71*(4), 1261–1286. https://doi.org/10.1093/bjps/axy058

Elgin, C. (1993). Understanding art and science. *Synthese, 95*, 13–28.

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

Fara, M. (2008). Masked abilities and compatibilism. *Mind, 117*(468), 843–865.

Faye, J. (2014). *The nature of scientific thinking: On interpretation, explanation and understanding*. Palgrave Macmillan.

Fridland, E. (2021). Skill and strategic control. *Synthese, Feb.* https://doi.org/10.1007/s11229-021-03053-3

Gibson, J. J. (1979). *The ecological approach to visual perception*. Psychology Press.

Ginet, C. (1980). The conditional analysis of freedom. In P. Van Inwagen (Ed.), *Time and cause: Essays presented to Richard Taylor* (pp. 171–186). Springer.

Grimm, S. R. (2006). Is understanding a species of knowledge? *British Journal for the Philosophy of Science, 57*, 515–536.

Hakli, R., & Mäkelä, P. (2019). Moral responsibility of robots and hybrid agents. *The Monist, 102*(2), 259–275. https://doi.org/10.1093/monist/onz009

Hannon, M. (2021). Recent work in the epistemology of understanding. *American Philosophical Quarterly, 58*(3), 269–290. https://doi.org/10.2307/48616060

Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.

Heras-Escribano, M. (2019). *The philosophy of affordances*. Springer International Publishing.

Hertzmann, A. (2020). Computers do not make art, people do. *Communications of the ACM, 63*(5), 45–48. https://doi.org/10.1145/3347092

Hetherington, S., & Lai, K. (2012). Practising to know: Practicalism and Confucian philosophy. *Philosophy, 87*(3), 375–393. https://doi.org/10.1017/S0031819112000289

Hills, A. (2016). Understanding why. *Noûs, 50*(4), 661–688.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Kelp, C. (2015). "Understanding Phenomena," Synthese, vol. 192, no. 12, pp. 3799-3816.

Khalifa, Kareem. 2012. "Inaugurating Understanding or Repackaging Explanation?" *Philosophy of Science*, vol. 79, no. 1, pp. 15-37.

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

Khalifa, K. (2013b). Is understanding explanatory or objectual? *Synthese, 190*(6), 1153–1171. https://doi.org/10.1007/s11229-011-9886-8

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge University Press.

Khosrowi, D., Finn, F., & Clark, E. (2023). Diffusing the creator: Attributing credit for generative AI outputs. In F. Rossi, S. Das, J. Davis, K. Firth-Butterfield, and A. John (Eds.), *AIES '23: Proceedings of the 2023 AAAI/ACM conference on AI, ethics, and society* (pp. 890–900). London.

Khosrowi, D., Finn, F., & Clark, E. (2024). Engaging the many-hands problem of generative-AI outputs: A framework for attributing credit. *AI and Ethics, March.* https://doi.org/10.1007/s43681-024-00440-7

Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World," in *Scientific Explanation,* ed. Philip Kitcher and Wesley Salmon (Minneapolis: University of Minnesota Press), pp. 410-505

Kieval, P. H. (Forthcoming). Representation learning without representationalism: A non-representational account of deep learning models in scientific practice. In J. Duran & G. Pozzi (Eds.), *Philosophy of Science for Machine Learning: Core Issues and New Perspectives*. Synthese Library

Kieval, P. H. & Westerblad, O. (Unpublished manuscript). *Deep Learning as Method-Learning: Pragmatic Understanding, Epistemic Strategies and Design-Rules*. Preprint available at https://philsci-archive.pitt.edu/23489/

King, R., Peter, O., & Courtney, P. (2023). Robot scientists: From Adam to Eve to genesis. In *Artificial intelligence in science*. Organisation for Economic Co-operation and Development (OECD). https://policycommons.net/artifacts/4595331/artificial-intelligence-in-sciencerobot-scientists/5419805/

Kneer, M., & Stuart, M. T. (2021). Playing the blame game with robots. In *Companion of the 2021 ACM/IEEE international conference on human-robot interaction* (pp. 407–11). HRI '21 Companion. Association for Computing Machinery. https://doi.org/10.1145/3434074.3447202.

Krenn, M., Kottmann, J., Tischler, N., & Aspuru-Guzik, A. (2020). Conceptual understanding through efficient inverse-design of quantum optical experiments. *arXiv.Org*, May. https://arxiv.org/abs/2005.06443v3.

Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., dos Passos Gomes, G., et al. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics, 4*(12), 761–769. https://doi.org/10.1038/s42254-022-00518-3

Kvanvig, Jonathan. 2003. *The Nature and Value of Knowledge* (Cambridge, UK: Cambridge University Press).

Lai, K. L. (2012). Knowing to act in the moment: Examples from Confucius' analects. *Asian Philosophy, 22*(4), 347–364. https://doi.org/10.1080/09552367.2012.729324

Le Bihan, S. (2017). Enlightening falsehoods: A modal view of scientific understanding. In S. Grimm, C. Baumberger, and S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 111–136). Routledge. https://www.academia.edu/21804009/Enlightening_falsehoods_a_modal_view_of_scientific_understanding

Lenhard, J. (2006). Surprised by a nanowire: Simulation, control, and understanding. *Philosophy of Science*, *73*(5), 605–616. https://doi.org/10.1086/518330

Lenhard, J. (2009). The great deluge: Simulation modeling and scientific understanding. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives* (pp. 169–188).

Lenhard, J. (2019). *Calculated surprises: A philosophy of computer simulation. Oxford Studies in Philosophy of Science*. Oxford: Oxford University Press.

Leonelli, S. (2009). Understanding in biology: The impure nature of biological knowledge. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives, 43–63* (pp. 189–209). University of Pittsburgh Press.

Leveringhaus, A. (2016). Ethics and Autonomous Weapons. *Palgrave Macmillan UK*. https://doi.org/10.1057/978-1-137-52361-7

Leveringhaus, A. (2018). What's so bad about killer robots? *Journal of Applied Philosophy, 35*(2), 341–358. https://doi.org/10.1111/japp.12200

Li, Y., Ni, Y., Croft, R. A. C., Di Matteo, T., Bird, S., & Feng, Yu. (2021). AI-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Sciences, 118*(19), e2022038118. https://doi.org/10.1073/pnas.2022038118

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.

McCormack, J., Hutchings, P., Gifford, T., Yee-King, M., Llano, M. T., & D'inverno, M. (2020). Design considerations for real-time collaboration with creative artificial intelligence. *Organised Sound, 25*(1), 41–52. https://doi.org/10.1017/S1355771819000451

Medeiros, L., Psaltis, D., Lauer, T. R., & Özel, F. (2023). The image of the M87 black hole reconstructed with PRIMO. *The Astrophysical Journal Letters, 947*(1), L7. https://doi.org/10.3847/2041-8213/acc32d

Mele, A. (2010). Moral responsibility for actions: Epistemic and freedom conditions. *Philosophical Explorations, June*. https://doi.org/10.1080/13869790903494556

Meskhidze, H. (2023). Can machine learning provide understanding? How cosmologists use machine learning to understand observations of the universe. *Erkenntnis, 88*(5), 1895–1909. https://doi.org/10.1007/s10670-021-00434-5

Miller, M., Nave, K., Deane, G., & Clark, A. (2020). The value of uncertainty. *Aeon* (blog). https://aeon.co/essays/use-uncertainty-to-leverage-the-power-of-your-predictive-brain.

Nersessian, N. J. (2022). *Interdisciplinarity in the making: Models and methods in frontier science*. MIT Press. https://doi.org/10.7551/mitpress/14667.001.0001

Nyholm, S. (2018). Attributing agency to automated systems: reflections on human-robot collaborations and responsibility-loci. *Science and Engineering Ethics, 24*(4), 1201–1219. https://doi.org/10.1007/s11948-017-9943-x

Pavese, C. (2024, Jan). The epistemology of skills. *Blackwell companion to epistemology*. https://www.academia.edu/114065364/The_Epistemology_of_Skills

Poliseli, L. (2020). Emergence of scientific understanding in real-time ecological research practice. *History and Philosophy of the Life Sciences, 42*(4), 51. https://doi.org/10.1007/s40656-020-00338-7

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.
Pritchard, Duncan. 2010. "Knowledge and Understanding." In *The Nature and Value of Knowledge: Three Investigations*, by Alan Millar, Adrian Haddock, and Duncan Pritchard. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199586264.001.0001/acprof-9780199586264
Räz, T., & Beisbart, C. (2024). The importance of understanding deep learning. *Erkenntnis, 89*(5), 1823–1840. https://doi.org/10.1007/s10670-022-00605-y
Rudy-Hiller, F. (2022). The epistemic condition for moral responsibility. In E. N. Zalta and U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/moral-responsibility-epistemic/.
Shech, E. (2022). Scientific understanding in the Aharonov-Bohm effect. *Theoria, 88*(5), 943–971. https://doi.org/10.1111/theo.12409
Shevlin, H., & Halina, M. (2019). Apply rich psychological terms in AI with care. *Nature Machine Intelligence, 1*(4), 165–167. https://doi.org/10.1038/s42256-019-0039-y
Smith, M. (2003). Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In S. Stroud and C. Tappolet (Eds.), *Weakness of will and practical irrationality*. Oxford University Press. https://doi.org/10.1093/0199257361.003.0002
Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy, 24*(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x
Stokes, D. (2021). On perceptual expertise. *Mind & Language, 36*(2), 241–263. https://doi.org/10.1111/mila.12270
Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A, 44*(3), 510–515. https://doi.org/10.1016/j.shpsa.2012.12.005
Strevens, M. (2024). Grasp and scientific understanding: A recognition account. *Philosophical Studies, 181*(4), 741–762. https://doi.org/10.1007/s11098-024-02121-x
Stuart, M. T., and Kneer, M. (2021). Guilty artificial minds: folk attributions of mens rea and culpability to artificially intelligent agents. *Proceedings of the ACM on Human-Computer Interaction, 5* (CSCW2: 363), 1–27. https://doi.org/10.1145/3479507
Stuart, M. T. (2016). Taming theory with thought experiments: understanding and scientific progress. *Studies in History and Philosophy of Science Part A, 58*(August), 24–33. https://doi.org/10.1016/j.shpsa.2016.04.002
Stuart, M. T. (2018). *How thought experiments increase understanding*. Routledge. https://doi.org/10.4324/9781315175027.ch30
Stuart, M. T. (2022). Sharpening the tools of imagination. *Synthese, 200*(6), 451. https://doi.org/10.1007/s11229-022-03939-w
Stuart, M. T., & Nersessian, N. J. (2019). Peeking inside the black box: A new kind of scientific visualization. *Minds and Machines, 29*(1), 87–107. https://doi.org/10.1007/s11023-018-9484-3
Sullivan, E. (2023). Do machine learning models represent their targets? *Philosophy of Science*, Oct, 1–11. https://doi.org/10.1017/psa.2023.151
Sullivan, E. (2022a). How values shape the machine learning opacity problem. In I. Lawler, K. Khalifa, & E. Shech (Eds.), *Scientific understanding and representation* (pp. 306–322). Routledge.
Sullivan, E. (2022b). *Link uncertainty, implementation, and ML opacity: a reply to Tamir and Shech*. Routledge.
Sullivan, E. (2022c). Understanding from machine learning models. *The British Journal for the Philosophy of Science, 73*(1), 109–133. https://doi.org/10.1093/bjps/axz035
Tamir, M., & Shech, E. (2022). *Understanding from deep learning models in context*. Routledge.
Tamir, M., & Shech, E. (2023). Machine understanding and deep learning representation. *Synthese, 201*(2), 51. https://doi.org/10.1007/s11229-022-03999-y
Vetter, Barbara. 2015. *Potentiality: From dispositions to modality*. Oxford Philosophical Monographs. Oxford University Press.
Vetter, B. & Schoonen, T. (Forthcoming). *The Epistemology of Ability*. Oxford: Oxford University Press.
Vihvelin, K. (2013). *Causes, laws, and free will: Why determinism doesn't matter*. Oxford University Press.
Westerblad, O. (Unpublished manuscript). *Scientific methods as a source of pragmatic understanding*.
Westerblad, O. (2023). *Making Sense of Understanding: A Pragmatist Account of Scientific Understanding*. PhD Thesis, University of Cambridge. https://doi.org/10.17863/CAM.108244
Wilkenfeld, D. A. (2017). MUDdy Understanding. *Synthese, 194*(4), 1273–1293. https://doi.org/10.1007/s11229-015-0992-x

Wilkenfeld, D. A. (2019). Understanding as compression. *Philosophical Studies, 176*(10), 2807–2831. https://doi.org/10.1007/s11098-018-1152-1

Zagzebski, L. (2008). *On epistemology*. Cengage Learning.