Scientific Understanding as Narrative Intelligibility

*Forthcoming in *Philosophical Studies*. Please cite published version when possible.

Gabriel Siegel

Washington University in St. Louis

Keywords

scientific explanation; interventionism; understanding; models; narratives; mechanisms

Abstract

When does a model explain? When does it promote understanding? A dominant approach to scientific explanation is the interventionist view (Woodward 2003). According to this view, when X explains Y, intervening on X can produce, prevent or alter Y in some predictable way. In this paper, I argue for two claims. First, I reject a position that many interventionist theorists endorse. This position is that to *explain* some phenomenon by providing a model is also to *understand* that phenomenon (Woodward 2003, Ylikoski and Kuorikoski 2010, Kuorikoski and Ylikoski 2015). While endorsing the interventionist view, I argue that explaining and understanding are distinct scientific achievements. Second, I defend a novel theory of scientific understanding. According to this view, when some model M promotes understanding, M makes available a distinctive mental state. This state is of the same psychological kind as when we grasp events in a narrative as bearing on some ultimate conclusion. To conclude, I show that, given this view, mechanistic explanations often provide a powerful source of understanding that many causal-historical models lack. This paper will be of interest to both philosophers of science and epistemologists engaged in the topics of explanation and understanding.

Introduction

Until recently, the topic of scientific understanding had gone somewhat into the background. There is now a resurgence of interest in the topic, e.g., Khalifa (2012, 2013, 2017), Strevens (2013), Kuorikoski and Ylikoski (2015), and Levy (2020). Ignoring the issue of understanding contrasts

with a long tradition in the scientific explanation literature. For example, Kitcher (1981) argued that theories of scientific explanation should "show us *how* scientific explanation advances our understanding" (1981: 508). Friedman said that "What is it about…scientific explanations...that gives us understanding of the world—what is it for a phenomenon to be scientifically understandable?" (1974: 5). Friedman later claimed that accounting for understanding is "the central problem of scientific explanation" (1988: 168).

A theory-neutral description of scientific explanation is that a model is explanatory when it contains information that is explanatorily relevant to some explanandum phenomenon. Many have disputed what counts as 'explanatorily relevant' (for a historical overview, see Salmon 1989). A dominant approach, which I adopt, is the interventionist view (Woodward 2003). According to this view, when X explains Y, intervening on X can produce, prevent or alter Y in some predictable way. Thus, on the interventionist conception, explanatory models represent counterfactual dependency information, e.g., that Y is counterfactually dependent on X.

In this paper I argue for two claims. First, I reject a position many interventionists endorse. This is that to *explain* some phenomenon by providing a model is also to *understand* that phenomenon (Woodward 2003, Ylikoski and Kuorikoski 2010, Kuorikoski and Ylikoski 2015).[1] Outside of interventionism, several philosophers have suggested that understanding reduces to explanation (see Khalifa 2012, 2017). By contrast, I argue that explaining and understanding are distinct scientific achievements. I suggest that for a model to explain P is a necessary, but *not sufficient*, condition for the model to promote understanding of P.

Second, I defend a novel theory of scientific understanding:

> **The Narrative Intelligibility View of Scientific Understanding**: some model M promotes *understanding* of some phenomenon P for some agent A iff (1) M explains P and (2) the explanatory relations between M's constituents and P are mentally represented by A in the *bearing on* manner of representation.

To mentally represent an explanatory relation between model constituents and some explanandum phenomenon, in the 'bearing on' manner of representation, is constitutive of the same kind of psychological state as when we grasp events in a narrative as furthering along, or leading to, some

[1] Khalifa (2012) also reads Grimm (2010) in this way.

ultimate conclusion. When M makes such a mental state available, I say that M exhibits 'narrative intelligibility.'[2]

The structure of the paper is as follows. In Section 1, I describe the interventionist view of scientific explanation. I describe how interventionists have discussed the relationship between explanation and understanding. In Section 2, I set up a counterexample to the claim that explaining and understanding are identical achievements. I describe a set of explanatory models and show that some are intuitively deficient in comparison to others. I argue that this intuitive difference cannot be accounted for by what theorists have called the model's explanatory 'force' or 'depth' (Hitchcock and Woodward 2003, Ylikoski and Kuorikoski 2010). Ultimately, I diagnose the intuitive deficiency as a matter of not promoting understanding.

In Section 3, I develop and defend the narrative intelligibility account of scientific understanding. I show that this view captures the contrastive deficiencies described in Section 2. In Section 4, I further motivate the proposed view by applying it to examples in chemistry, psychology and biology. By looking at these illustrative cases, I emphasize that mechanistic models often provide a powerful source of narrative intelligibility. In many cases, the difference between a model that promotes understanding, and one that does not, depends on whether the model provides mechanistic information about how the explanandum phenomenon is produced. This upshot further motivates the importance of recognizing mechanistic explanations in addition to causal-historical explanations (see Salmon 1984, Craver 2007, Siegel and Craver 2024).

1 Interventionism, Explanation and Understanding

Interventionism is the view that models explain by providing counterfactual dependency information (Woodward 2003). According to this view, when X explains Y, ideal interventions on X, given the right background conditions, result in predictable manipulations of Y. In this way, explanatory models represent generalizations that hold between variables X (the explanans) and Y (the explanandum) given a certain range of interventions to X. Such generalizations provide counterfactual dependency information and answer a certain range of *what-if-thing-had-been-different* questions, i.e., "w-questions" (Woodward 2003).

[2] This paper is concerned specifically with *scientific* understanding. I leave it open-ended how the proposed account generalizes to other forms of understanding.

It is common for interventionists to assert that explaining and understanding are the same scientific achievement. For example, Woodward says that "The account I defend takes this and other explanations to provide understanding by exhibiting a pattern of counter-factual dependence between explanans and explanandum" (2003: 13). Representing counterfactual dependency information is, for Woodward, the signature of a model's explanatory status. Thus, on a plausible reading of Woodward, providing an explanatory model also means that the model promotes understanding.

Interventionists Ylikoski and Kuorikoski endorse the target view more explicitly. They say that "in our view, the fundamental criterion according to which understanding is attributed is the ability to make inferences to counterfactual situations, the ability to answer contrastive *what-if-things-had-been-different* questions (*what if*-questions) relating possible values of the explanans variables to possible values of the explanandum variable" (2010: 205). They say that "explanation and understanding are not different kinds of activities, mental processes or methods. Successful explanations convey the ability to provide answers to contrastive what-if questions, and understanding is based on a body of knowledge that provides the basis for these answers" (Kuorikoski and Ylikoski 2015: 3820). Thus, the criterion Kuorikoski and Ylikoski employ to determine whether some model M is explanatory, i.e., that it contains accurate counterfactual dependency information, is the same criterion for M to promote understanding. Under such an account, there is no room for a model that is explanatory but fails to promote understanding. In the proceeding sections, I argue that this is a mistake.

2 Case Study: Lewis's Letter

In this section, I develop a counterexample to the claim that explaining and understanding are identical achievements. I outline a set of models where a proper subset is intuitively deficient. I show that explanatory depth, implicated by interventionist theorists such as Woodward, Hitchcock, Ylikoski and Kuorikoski, cannot adequately capture the difference. Ultimately, I diagnose the contrastive deficiency as a matter of not promoting understanding. I suggest that while all models in the set are explanatory, only one provides understanding. This shows that a model can be explanatory yet fail to promote understanding. The models I discuss here, i.e., statements about events in a causal history, are best construed as *etiological models*. Etiological models represent some event or output state and reveal its antecedent causes. Such models represent counterfactual

dependency relations or more broadly just provide information regarding how some effect was caused. In Section 4, I focus on cases of *constitutive* explanations in chemistry, psychology and biology.[3]

The example I discuss here comes from David Lewis (1986b). In this case, Lewis writes a letter of recommendation for X. X gets a professorship position to which they applied and wouldn't have gotten the job if Lewis hadn't written the letter. While employed at the job, X meets and marries their partner, something that would not have happened had they not taken the job. Consequently, they have children, and their children have children. Now, consider the following statement on its own (i.e., without the background information just described):

(1) Lewis writing the letter caused X's grandchildren to exist.

As Woodward notes, "whether or not we accept this judgment, virtually everyone will agree that there is something non-standard, or misleading about [the statement]" (2010: 292). How can we explain this intuition that there is something deficient about (1), understood as a model of this causal history? The reason cannot be that (1) is not explanatory. It meets the minimal interventionist standard for being explanatory since it provides counterfactual dependency information. It answers w-questions, i.e., it tells us that if Lewis had not written the letter, then X's grandchildren would not exist.

Now consider the following statement:

(2) Lewis writing the letter caused X to get a job at Cornell. X met their partner while teaching at Cornell. They had children. Their children had children.

By contrast to (1), (2) seems to be a saliently less deficient model representing that writing Lewis's letter causally explains the birth of X's grandchildren.

What is the difference between (1) and (2) such that the latter is intuitively less deficient as the former? (1) represents less of the causal history relevant to the explanandum (i.e., the birth of X's grandchildren) than (2). But the relevant difference cannot consist in the amount of the causal history the statement represents. For example, consider:

[3] Constitutive models, as distinct from etiological models, tell us how some phenomenon is produced by the organization of entities and their activities (Machamer et al. 2000, Craver 2007).

> (3) Lewis writing the letter caused X's grandchildren to exist. Lewis wrote the letter because Lewis was impressed by X's dissertation. Lewis sent the letter to Cornell via USPS mail. Y has great admiration for Lewis's work. Y was on the search committee that hired X.

This statement provides as much causal information as (2). The additional events add further counterfactual dependency information omitted in (1) that meets the interventionist criteria for explanatory relevancy. For example, if one intervened on the transport of the letter, this would plausibly disrupt the birth of X's grandchildren. We could intervene on USPS's transportation of the letter by stealing it from the mail truck and putting it in a shredder. Nevertheless, (3) remains intuitively deficient in the same way that (1) is. It is unclear, without the right background knowledge, how the cluster of causal information in (3) leads to the birth of X's grandchildren. Hence, the relevant difference cannot merely be the *amount* of explanatorily relevant causal information that the two statements provide. In other words, the *number* of w-questions the statements answer cannot account for the difference. Both (2) and (3) answer the same number of w-questions about the explanandum, but (3) remains intuitively deficient compared to (2).

In this vein, Kuorikoski and Ylikoski might suggest that (2) promotes a greater amount of what they call "understanding" than (1), and that this captures the difference. As they say, "the number and precision of correct what-if inferences determines how much one understands" (2015: 3819). However, since statements (2) and (3) enable the same number of what-if inferences, this cannot account for the intuitive contrastive deficiencies between models.

Woodward (2010) follows Lewis (1986b) in saying that the issue with (1) is *stability*. The counterfactual dependence relation specified in (1) is not stable. If the background conditions were only slightly different, e.g., if X's partner didn't take a job at the same university, the counterfactual relation would not continue to hold (2010: 292). While this might explain (1)'s general misleadingness, it cannot explain its being *contrastively deficient* with respect to the other statements. Notice that stability is equally an issue with (2). Any of these minor changes would equally disrupt the counterfactual dependency relation specified in (2). This is because all statements (1) – (3) represent the same causal history. If the causal connection between Lewis's letter and X's grandchildren is unstable, then *any* representation of the causal history in between

will also be unstable. Hence, a lack of stability cannot be the intuitive difference in model deficiency that we are after.

Outside of the number of w-questions the statements answer, interventionists may argue that the explanatory 'depth' or 'force' of (2) is greater compared to (1) and (3). In this vein, they might suggest that the notion of 'explanatory depth' can explain why (2) is intuitively less deficient than (1) and (3).

Hitchcock and Woodward (2003) were plausibly the first to attempt a systematic account of the notion of 'explanatory depth.' Their account builds off the interventionist conception of explanation. For them, the explanatory depth of some model M is a function of how *invariant* the generalization represented in M is. Invariance is understood as the range of alterations to the explanans variable such that the generalized counterfactual relationship continues to hold. Some generalization might be more or less invariant for a number of reasons. For example, a generalization G^* might provide more accurate values for the explanandum variable than G. Furthermore, G^* might be invariant under a greater range of interventions onto the explanans variable than G. Hitchcock and Woodward also note that G might be more sensitive to background conditions than G^*, i.e., it might be less stable (see Hitchcock and Woodward 2003: 184-185, 187).

Focusing for the moment on (2) and (3), these two statements are not clearly different regarding invariance. While invariance is thought to be a feature of *generalizations*, and (2) and (3) involve a particular causal history, we can investigate the extent to which invariance might capture the intuitive difference between the statements. Like stability, invariance is a feature of causal structures, not the *representations* of causal structures. Again, an alteration of invariance means that the range of interventions to the explanans variables, such that the counterfactual relationship continues to hold, will be different. However, for (1) - (3), the causal history is the same. Like for stability, representing different parts of the same causal history won't alter invariance. For example, by representing USPS delivery in (3), but not (2), this does not change the range under which interventions to Lewis writing the letter will alter the birth of X's grandchildren. Furthermore, disruption of the USPS delivery, an intermediary causally relevant factor, will equally disrupt the counterfactual relationships represented in (2), even if USPS delivery is not represented in (2). Again, the reason is that there is a difference in representation, not the causal history that is represented. Thus, invariance under a greater number of interventions, for example, cannot occur for (2) as opposed to (3).

Ylikoski and Kuorikoski (2010) provide five dimensions of explanatory depth. One might also appeal to these as an account of the difference between (1), (2), and (3). However, I argue that they cannot.

The first dimension, *non-sensitivity*, we have already dealt with above. This is the same notion as stability. The second dimension is how *precisely* the explanandum is specified in relation to a contrast space (2010: 2010-211). Since the characterization of the explanandum is the same in all three statements, i.e., "the birth of X's grandchildren," this cannot account for the difference. Their third dimension, *factual accuracy*, is relevant in the context of idealization or abstraction in models. This is also not relevant here, as these models involve representations of singular events. No abstraction is involved. Fourth is *degree of integration* for some model into existing explanatory knowledge. Such integration, the authors reason, will consequence in the capacity to answer a greater range of w-questions (2010: 213). However, by assumption, in this case we do not have any other background information about Lewis, X, X's partner, X's children, etc. Thus, the idea that one model would allow for greater integration with other relevant knowledge is not applicable here either.

Lastly, the authors discuss what they call "cognitive salience." They say, "cognitive salience refers to the ease with which the reasoning behind the explanation can be followed, how easily the implications of the explanation can be seen and how easy it is to evaluate the scope of the explanation" (2010: 214). This is on the right track, but we need to look more closely at why the reasoning behind certain explanations might be easier to follow than others. An attractive explanation is the narrative intelligibly account developed shortly. However, unlike Ylikoski and Kuorikoski, I don't construe this as part of an explanation's depth. On my view, a model's explanatory depth depends solely on the nature of the counterfactual dependency information it represents. It does not depend on any kind of mental state the explanation makes available.[4]

Thus far, I've argued that interventionists lack the resources to capture (2)'s intuitive superiority in terms of *explanatory* superiority, e.g., in terms of its explanatory "power" or "depth." In other words, on the assumption that we should adopt the interventionist account of explanation, (2) provides the same degree of explanatory power as (3).

[4] This is inspired by the *ontic view* of explanation, which for simplicity I've avoided discussion of in the present article. On the ontic view, explanations are causal structures in the world within which explanandum phenomena are situated (Salmon 1984). Explanations are contrasted with explanatory models, i.e., representations of such causal structures. Some interventionists are also ontic theorists (e.g., see Craver 2007).

Readers might respond to this result in different ways. For example, interventionists might acknowledge the need to analyze models in a way that transcends their explanatory quality. In subsequent sections, I follow this route and argue that (2) promotes understanding while (1) and (3) do not. However, other readers might, by contrast, view this result as a strike against interventionism. Rather than incorporating a distinct notion of 'understanding' into the interventionist framework, perhaps non-interventionist theories of causal explanation can be adopted that capture (2)'s superiority in terms of explanatory quality. While space prevents an exhaustive review,[5] there are reasons for skepticism that this is the best approach. First, there are independent motivations for the interventionist theory of explanation, including its response to notorious difficulties with previous accounts, such as Hempel's (1965) covering law and Kitcher's (1981) unificationist account. Furthermore, interventionism has been shown to capture explanatory practices across a wide range of scientific disciplines. I will not recapitulate these points here (see e.g., Woodward 2003, 2010, Craver 2007, Rescorla 2018).

Second, many non-interventionist frameworks don't seem capable of capturing the intuitive difference either. For example, consider Lewis's (1973, 1986a) counterfactual view of causal explanation. On this view, like interventionism, a statement is explanatory when it provides counterfactual dependency information. However, for Lewis, claims about counterfactual dependency relations are not understood as claims about possible interventions. Rather, a counterfactual dependency claim, e.g., "if A were the case, then B would be the case," is true iff all the possible worlds in which A and B occur are "closer" (by virtue of measures that needn't detain us here) to the actual world than possible worlds in which A occurs and B doesn't. Again, since (1) – (3) represent the same causal history, a possible worlds account cannot capture the difference.

Furthermore, it might be postulated that causal-mechanical theories such as Salmon (1984, 1998) and Dowe (2000) might be better positioned to capture (2)'s superiority in terms of explanatory quality. For example, it might be held that (2) provides information about actual *causal processes* that link explanans to explanandum in an unbroken chain of intersections, e.g.,

[5] For example, I will not describe how Hempel's (1965) covering law account might capture the intuitive differences in terms of explanatory quality. Perhaps certain statements allow for chains of covering laws from the initial event to the explanandum, while other statements do not. However, it is precisely this sort of treatment, where covering laws are applied at each step of a temporally ordered sequence of events, that is implausible when applied to other cases, e.g., cows lying down before a storm (see Scriven 1962). Given this and other well-established difficulties with Hempel's account (e.g., see also Salmon 1989), I do not entertain this possibility here.

via "mark transmissions" or "exchanges of conserved quantities," where (1) and (3) do not. Since none of the statements (1) – (3) describe such causal processes, this response also appears implausible.

Lastly, consider mechanistic explanation (which I return to in greater detail in Section 4). Mechanistic models explain some phenomenon by showing how that phenomenon is the behavior or property of a mechanism. A mechanism is composed of component parts and the organized activities of those parts. As one reviewer suggests, perhaps (1) – (3) can be construed as models of mechanisms. The causal histories represented in (1) – (3) can be decomposed into organized parts that interact with one another. In this vein, we can reconceptualize (1) – (3) as models of historical mechanisms. Glennan (2010) calls such mechanisms "ephemeral mechanisms." Perhaps the intuitive superiority of (2) is a result of providing a better sketch[6] of the entire historical mechanism while (3) only represents components of its first stage. More broadly, the information provided in (2) is more temporally spread out than (3) and perhaps this explains its intuitive superiority. However, this does not seem to capture the relevant difference. To illustrate, consider the following statement on its own:

> (4) Lewis writing the letter caused X to get a teaching job. X met Y in 1991. A and B were born. The grandchildren of X were born.

Here, each component of (4) refers to the *same components* (i.e., events) of (2). Those components are simply redescribed. Thus, the component parts represented in (2) and (4) are equally temporally distributed between the explanans and the explanandum. However, (4) is still intuitively deficient relative to (2). The events are described in such a way that, without the right background knowledge, it is unclear how one event leads into another.

In this section, I've shown that there is something deficient about the statements (1), (3) and (4) as compared to (2). The notion of explanatory depth, as discussed by theorists such as Hitchcock, Woodward, Ylikoski and Kuorikoski, I've shown, cannot account for the deficiency. Furthermore, possible worlds, causal-mechanical and mechanistic[7] theories of explanation alone seem ill-equipped to explain the intuitive difference we are after in terms of explanatory quality.

---

[6] Mechanism *sketches* are incomplete descriptions of mechanisms (see Machamer et al., 2000, Craver 2007, Craver and Darden 2013).

[7] For discussions of how interventionist, causal-mechanical and mechanistic approaches relate to one another, see Craver (2007) and Siegel and Craver (2024).

Ultimately, in Section 3, I'll show that (2) promotes understanding while (1), (3) and (4) do not. Since (1) – (4) are explanatory, but only (2) promotes understanding, this will show that explaining and understanding are distinct achievements.

3 Scientific Understanding and Narrative Intelligibility

The central question of this paper is: are there different conditions under which a model explains and under which it promotes understanding? The distinction between a merely explanatory model, and a model that provides understanding, bears important similarities to the distinction between causal knowledge and understanding made by some epistemologists.[8] Consider a representative passage by Kvanvig (2003):

> Understanding requires, and knowledge does not, an internal grasping or appreciation of how the various elements in a body of information are related to each other in terms of explanatory, logical, probabilistic, and other kinds of relations that coherentists have thought constitutive of justification (2003: 192-193).

For Kvanvig, to know a set of propositions P^n doesn't imply that one understands the set. The distinct cognitive state of understanding P^n requires "grasping" certain kinds of relations between the propositions within P^n. The idea that understanding consists in 'grasping' relations among propositions in a body of information is held by many, including Elgin (2007), Khalifa (2013) and Hills (2016). While the concept of 'grasping' often remains opaque, some have offered systematic accounts, e.g., Khalifa (2013) and Hills (2016). I discuss these views in greater detail below.[9]

I now turn to scientific understanding. Like theorists such as Kvanvig, Pritchard, Khalifa and Hills, I assume that the key to understanding is a kind of cognitive 'grasping' of certain

[8] This similarity will be helpful to note even though I'm concerned specifically with *scientific* understanding, plausibly a proper subset of understanding as discussed more broadly.

[9] Arguments that knowledge and understanding are distinct epistemic states come in a variety of forms. For example, Pritchard (2014) argues that understanding constitutes a distinct kind of cognitive achievement from knowledge (also see Elgin 2007). In addition, many have argued that understanding can involve cases of 'environmental luck,' while knowledge cannot (e.g., Kvanvig 2003, Pritchard 2014, Hills 2016). An example of environmental luck is walking into a library full of history books with false information, pulling out the one accurate book, reading it and believing what you read. Due to the etiology of belief formation in such cases, these are considered cases of understanding but not knowledge (Kvanvig 2003). Furthermore, while knowledge can plausibly be transmitted through testimony, understanding cannot (Hills 2016). Understanding also plausibly comes in degrees while knowledge does not (Kvanvig 2003). For some arguments against the distinction between knowledge and understanding, see Lipton (2009) and Grimm (2014). I do not enter the debate regarding these contrasting features of knowledge and understanding. My focus is on understanding and explanation in scientific contexts, and what conditions need to be met in order to achieve scientific understanding, as opposed to mere explanation.

relations between pieces of information. In the case of scientific understanding, these are causal explanatory relationships between constituents of scientific models. In the present section, I develop an account of grasping and show its relationship to interventionist explanations. In Section 4, I apply the view to various scientific contexts. Before these developments, I describe some preliminary assumptions.

First, I adopt the assumption, held by many epistemologists, that understanding is *factive* (Kvanvig 2003: 190, also see Kuorikoski and Ylikoski 2015).[10] In scientific contexts, this means that, adopting the interventionist view of explanation, to understand an explanandum phenomenon P that some model M explains, M must accurately represent how P is counterfactually dependent on certain variables. In other words, M must be explanatory.[11]

The assumption that understanding is factive also distinguishes understanding from the mere *sense of understanding*. This is the mere feeling, or phenomenological experience, of understanding some phenomenon. Trout (2002) persuasively argues that having a psychological sense of understanding is misleading and not an accurate indicator of truth (also see Rozenblit and Keil 2002). The factive nature of understanding also provides an answer to Friedman (1974) and Kitcher's (1981) question regarding *how* scientific explanations promote understanding. If understanding is factive, then explanations are necessary conditions for scientific phenomenon to be understood. As I suggest, the further step toward understanding involves the appropriate grasping of explanatory relations between constituents of models.

Second, in motivating the importance of scientific understanding, I am not advocating a *cognitive account* of scientific explanation (Churchland 1989, Bechtel and Abrahamsen 2005). In contrast to cognitive views, I assume that what makes some model M explanatory is not dependent on whether M promotes any kind of cognitive state, understanding or otherwise. Following the interventionist conception, I assume that M is explanatory iff it contains accurate counterfactual dependency information regarding an explanandum phenomenon and some explanans.[12]

---

[10] For a critique of this position, see Elgin (2007).

[11] This view is also held by Strevens (2013). I take the facticity of understanding to mean that idealized models, i.e., models that falsely represent causal structures to make them more intelligible, do not promote understanding. Alternatively, one might construe idealized models as intentionally omitting certain explanatory features in order to engender understanding. If construed in this way, idealized models can be explanatory. In any case, idealized models raise a set of complicated questions that I needn't consider here.

[12] A third relatively orthogonal point is that many have argued that knowledge is not necessary for understanding (Kvanvig 2003, Hill 2016). As mentioned in a previous footnote, theorists point to cases of so-called 'environment luck' to show this. However, the same cannot be said for the relationship between a scientific model being explanatory and it promoting understanding. I assume that a model being explanatory is a necessary condition for it to promote

I now develop the proposed account. On my view, what does it mean to *grasp* the relationship between constituents of some model M, such that one *understands* the phenomenon that M explains? In other words, what kind of grasping is involved when M makes some phenomenon *intelligible*? I suggest that the relevant grasping, at least in scientific contexts, is constituted by a particular way of mentally representing causal relations. I call this way of representing the *bearing on* manner of representation. Paradigmatic instances of such manners of representation occur in the context of reading narratives. For this reason, I begin to develop this notion by discussing narrative explanations. When we grasp relations between events in narratives, our mental states represent causal relations in the appropriate way as to promote understanding.

What are narrative explanations? Goudge (1961) suggested that narrative explanations situate an event within the context of other events that lead up to it. However, narrative explanations do not merely situate some event within the context of others. As Velleman says, "a story does more than recount events; it recounts events in a way that renders them intelligible, thus conveying not just information but also understanding…what makes a story good…is its excellence at a particular way of organizing events into an intelligible whole" (2003: 1). How do narratives make a sequence of events intelligible?

As Mink (1970) and Vellman (2003) note, narratives link events in a sequence to some outcome or end-state. In virtue of linking events to this outcome, the events are understood as a cohesive unit constituting a story. For example, in the tale of Treasure Island, Velleman notes that, "each major event can be regarded as either motiving or furthering or hindering or somehow bearing on the pursuit of Flint's treasure" (2003: 9). The issue is not merely whether the events in the sequence causally explain the outcome, e.g., in Woodward's interventionist sense. Rather, it, "depends on how well the events in the story can be *grasped together as bearing on this outcome* in some way or other" (my italics, 2003: 10). Vellman's insight is that, much like in the discussion of understanding had by epistemologists, narratives do something more than outline a set of

understanding. This is because a model's explanatory status, given the interventionist conception, only depends on whether it accurately contains at least some counterfactual dependency information. In this way, whether a model is explanatory does not depend on the process through which a model is constructed or how the counterfactual relations are discovered. A model's explanatory status depends solely on whether it contains explanatorily relevant information. Thus, many counterexamples, such as environmental luck cases, that purportedly show that knowledge is not necessary for understanding, do not carry over to show that a scientific model's explanatory status is not necessary for it to provide understanding.

causally related events. They present the causal relations between those events in such a way that make them intelligible.

The relevant 'grasping' of the relations between events in narratives, as Vellman notes, is how the events "bear on," or "further along," some ultimate conclusion. This grasping is what I call *bearing on* manners of representation. In other words, the bearing on manner of representation is a special way that mental states represent causal relations. When mental states represent causal relations in a bearing on manner between causal relata, i.e., two entities, events or variables counterfactually dependent on one another, this, I suggest, is the grasping of explanatory information constitutive of understanding. Representing causal relations in this way makes them intelligible.[13] By representing such relations in the bearing on manner, we comprehend *how* the causal relations hold between some relata. In other words, when we represent that C* causes C in the bearing on manner, we comprehend how C* causes C. In this vein, we have an articulation of grasping in scientific understanding. *Grasping* of explanatory information involves an agent's mental state representing explanatory causal relations in a bearing on manner.[14] These manners of representation are paradigmatically made available in narrative contexts.

When the mind represents that a set of events, variables, mechanistic components, etc., are causally relevant to some explanandum phenomenon, in the bearing on manner of representation, those constituents' contribution to bringing about the explanandum are intelligible. In such cases, the constituents of the model are mentally represented *in light of* that phenomenon and their roles in bringing about that phenomenon are perspicuous. The same goes for narratives. Each major event in some narrative, once the ending is revealed, can be viewed in light of that ending and acknowledged as playing a specific role in bringing about that conclusion. This intelligibility allows events and their outcome to be represented by the mind, in Vellman's words, as an "intelligible whole" (2003: 1).

[13] There are other accounts that construe understanding as involving a kind of intelligibility, e.g., de Regt (2009). Some have found such accounts unsatisfying, e.g., see Khalifa (2012).

[14] By "causal" explanatory relations, I mean to include organizational and constitutive relations that appear in mechanistic explanations (which I discuss in Section 4). These relations indicate that certain parts are components of some mechanism and organized in a particular way within that mechanism. There is disagreement on how constitutive relations (or 'constitutive relevance') is related to causal relations. I don't enter this debate here. For some recent discussions, see Baumgartner and Gebharter (2016), Baumgartner and Casini (2017), Prychitko (2019), and Craver et al. (2021). In any case, the crucial point here is that, on my view, representing organizational and constitutive relations, in the bearing on manner, plays an important role in understanding phenomena that mechanistic models explain.

To begin to illustrate the bearing on manner of representation, consider Cummins's (1975) account of functional analysis. On a plausible reading, narrative intelligibility is in the background of this account. Cummins advocated what he called "analytic accounts" of the capacity of some system. Roughly, this involved decomposing that system into subsystems, and showing how the capacities of each subsystem contribute to the overall capacity of the system.[15] For illustration, consider an assembly line that, taken together, has the capacity to produce some product. Cummins says, "Each point on the line is responsible for a certain task…Against this background, we may pick out a certain capacity of an individual…his function on the line is doing whatever it is that we appeal to in explaining the capacity of the line as a whole" (1975: 760). When providing an analytic account, a thing's function is specified only with respect to the role it plays in the behavior or output of a system. On this view, if it were unclear what role some assembly line task played in the line's end-state (i.e., its product), then it would be unclear what that task's function was. Within the context of analytic accounts, a task's function is thus only specified once it is viewed *in light of* some end-state or overall capacity of the system of which the task is a constituent.

Cummins's constraint on the specification of function is importantly analogous to how, under the narrative intelligibility view, models do or do not promote understanding. Consider a Cummins-style model of some system. The representation of a subsystem of that system would appear 'functionless' if it were unclear how the behavior of that subsystem contributes to the overall system's capacity. Likewise, consider the representation of an event in a causal history. If it were unclear how that event 'bears on' or 'furthers along' the causal history's end-state, then it would be unclear what role that event played in the history's end-state. Both instances exhibit a lack of understanding given the incapacity to represent causal relations between a model's constituent and the model's explanandum in a bearing on manner. To represent that some component of a model C causally explains some explanandum phenomenon P, in a bearing on manner, C must be viewed in light of P and it must be clear what role C played in bringing about P. Such bearing on manners of representation are made available in narratives and successful Cummins-style functional analyses.

The bearing on manner of representation, as understood here, is not identical to mind-independent causal relations. Bearing on is a *way of representing* causal explanatory relations.

[15] This analysis is perhaps most precise when the 'system' is a mechanism and the 'subsystems' are the component parts of that mechanism (see Craver 2001).

These ways of representing, as used in this technical sense, are only present in *states of mind*. Since the bearing on manner of representation is not an objective feature of the world, it does not have accuracy conditions. Bearing on is a fundamentally mental way of representing how events, mechanical parts, etc., causally relate to one another.

However, given the factive nature of understanding, mental states involving bearing on manners of representation *depend* on the existence of certain mind-independent causal relations in order to engender understanding. These causal relations have accuracy conditions, i.e., C* causes C in virtue of the world being such that C* does in fact cause C. Take some model M with two constituents C and C*, where C is represented as counterfactually dependent on C*. In order for M to promote understanding, this counterfactual relation must be a part of the world's causal structure. Thus, when some mental representation of a causal relation between C* and C constitutes understanding, it must be the case that C* causes C. However, we can mentally represent that C* causes C without representing this causal relation in a bearing on manner. Such differences in mental states are constitutive of the difference between explaining and understanding. Furthermore, we can represent that C* causes C in a bearing on manner of representation without it being the case that C* causes C. This constitutes the mere *sense of understanding* or *misunderstanding*, which involves non-factively representing causal explanatory relations. This should be distinguished from mental states that are both non-factive and fail to represent causal relations in a bearing on manner. We might call such states *non-understanding*.

I consider the bearing on manner of representation to be an irreducible fundamental form of mental representation. The bearing on manner is a primitive way in which minds make causal explanatory relations between worldly entities intelligible. Given its primitive nature, defining understanding stops at the bearing on manner of representation. However, to characterize the bearing on manner of representation, I've indicated that its paradigmatic instances occur in the contexts of narratives. Furthermore, I've indicated that when we represent that C causes P in a bearing on manner, C is viewed in light of P and it's clear what role C played in bringing about P. Mental states that represent causal relations in bearing on manners of representation can come in different forms. For example, they can be thoughts about events in narratives, mental representations of flowcharts, mental animations of interacting components of a mechanism, etc.

On the proposed account, whether M promotes understanding depends on facts about the mental states of individuals. This means that, in certain cases, causal relations can be represented

by two distinct agents, but only one might represent the causal relations in a bearing on manner. Furthermore, at two different times, one agent might fail to represent causal relations in a bearing on manner, but later come to represent them in a bearing on manner in an "aha!" moment. There are plausibly many reasons for such differential understanding. One is that the capacity to represent in a bearing on manner depends on an individual's background knowledge. Depending on the degree of background knowledge an agent A has about some explanandum phenomenon P, M will be more or less likely to promote understanding of P for A. For example, going back to the case of Lewis's letter, with sufficient knowledge about X's life, statement (1) might promote understanding for an agent A. For another agent, A^*, who lacks the requisite background knowledge, (1) will not promote understanding.

To sum up, my suggestion is that, in scientific models that promote understanding for some agent, that agent represents explanatory causal relations between the model's constituents and the explanandum phenomenon in a bearing on manner. Take some explanandum phenomenon P that some model M explains. The proposed account is as follows:

> **The Narrative Intelligibility View of Scientific Understanding**: some model M promotes *understanding* of some phenomenon P for some agent A iff (1) M explains P and (2) the explanatory relations between M's constituents and P are mentally represented by A in the *bearing on* manner of representation.

In this case, M's constituents are psychologically analogous to events in a story and P to the story's conclusion. To represent causal relations among a model M's constituents in a bearing on manner is to understand the phenomenon M explains. In this vein, mere explaining occurs when agents represent causal information in M that meet the interventionist criteria for explanatory relevancy. Understanding, in addition to explaining, occurs when agents represent the causal explanatory relations in a bearing on manner.[16]

This does not imply that, for a model to promote understanding, it must be a narrative, i.e., it must tell a story in the literal sense. Philosophers have discussed the use of narratives in science (e.g., see Goudge 1961, Morgan and Wise 2017). However, to promote understanding, the model only must make the same kind of psychological state available. What is important is the particular

[16] I assume that explanatory relations between two constituents of models can be represented in the 'bearing on' manner, which we might call 'pairwise' intelligibility, as well as for explanatory relations between multiple interacting constituents of a model, which we might call 'global' intelligibility.

psychological kind of which its paradigmatic instance occurs in the context of representing relations between events, characters and their activities in narratives.

In many cases, whether a model fosters mental representations of causal relations in a bearing on manner, and thus promotes understanding, is intuitive. To provide an illustration, let's return to the case of Lewis's letter. Consider again these two statements:

(1) Lewis writing the letter caused X's grandchildren to exist.

(2) Lewis writing the letter caused X to get a job at Cornell. X met their partner while teaching at Cornell. They had children. Their children had children.

(1) tells us that Lewis's letter of recommendation is causally relevant to the birth of X's grandchildren. Given the information provided in (1), however, the event cannot be comprehended as bearing on the outcome of the grandchildren's birth. This is an intuitive fact about the intelligibility of (1) as a representation of the causal history. The gaps in the causal history represented in (1) impede us from representing the causal relationship between Lewis's letter and X's grandchildren being born in a bearing on manner. In light of the birth of X's grandchildren, we cannot grasp the role Lewis's letter played in bringing that event about. The same goes for (4), which redescribed the events in (2) in such a way that made their causal relationships unintelligible.

These impediments are not present in (2). (2) represents the right bits of the causal history, in the right way, to enable understanding. In that case, we can comprehend how the writing of Lewis's letter (an event) leads to the birth of X's grandchildren (an end-state). Readers who find this difference salient are tracking the distinction between a mental state that represents causal relations in a bearing on manner and one that does not. Thus, in (2), but not in (1), we represent that Lewis's letter causes the birth of X's grandchildren in a bearing on manner. Given the proposed account, this demonstrates that (2) promotes understanding while (1) does not. Furthermore, in Section 2, I showed that statements (2) and (3) were equally explanatory, i.e., that they answered the same number of w-questions. However, (3) was intuitively deficient compared to (2). We are now able to diagnose this deficiency. (2) promotes understanding while (3) does not, even though they both contain the same amount of counterfactual dependency information.

This shows us two things. First, explaining and understanding are distinct achievements. (1) and (3) are explanatory yet they fail to promote understanding. Second, it provides a convincing

case for the narrative intelligibility account of understanding. Unlike the notion of explanatory depth, or other features of interventionist (and plausibly many non-interventionist) explanations, the proposed view can account for the intuitive contrastive deficiencies between the statements.

So far, my defense of the proposed view of scientific understanding has appealed to salient intuitions about cases and has been largely conceptual. This raises a legitimate worry expressed by Trout (2017). Trout says that, "the existing philosophical accounts of this fundamentally psychological notion—understanding—are not formulated in way that could be confirmed by scientific evidence. Instead, they characterize understanding by performing conceptual analysis on the experience of understanding…If we had a psychologically informed and empirically rigorous account of 'grasping', that would go a long way toward characterizing understanding. But we don't" (2017: 238-239). In response, the psychological capacities involved in representing causal relations in a bearing on manner is well-established by psychologists. As we'll now see, these capacities can be disrupted with brain damage and are underdeveloped in infants.

Consider a couple experimental paradigms that study the capacity to construct narratives for individuals with brain damage or disease (Ash et al. 2006, Keven et al. 2017). In these studies, subjects are provided with a picture book that tells a story. The story is about a boy who loses and later finds his lost pet frog. The authors say:

> Subjects were scored positively for global connectedness if they recognized that the frog found at end of the story is the same frog that figures in the opening pages, the one they have been searching for. Subjects were scored as not showing global connectedness if they talked about the event in which the boy and dog come upon their frog but did not indicate that the frog had been present earlier in the story (Ash et al. 2006: 1406-1407).

Here, "global connectedness" essentially requires that participants identify the outcome event (finding the frog) as a resolution of an earlier event (losing the same frog). In other words, as Keven et al. (2017) put it, this measure "assesses the overall point of the story, that the boy and his dog search for and find the escaped frog" (2017: 106). The participants must show how the initial event further develops, and is related to, the outcome event. This task can be construed as an attempt to operationalize, or track, the capacity to represent causal relations in a bearing on manner. Maintaining character identity across events is necessary for such manners of representation. In this way, showing disruption of global connectedness indicates disruption of the capacity to represent in a bearing on manner. Using this task, Ash et. al (2006) showed that this

capacity is disrupted in progressive aphasia and frontotemporal dementia. Keven et al. (2017) showed that this capacity is intact in episodic amnesia.

Keven (2016) refers to "narrative binding" as the capacity to bind events together into a temporal and causal order. As Keven notes, narrative binding appears to be absent in infants, where they have what is called "childhood amnesia" (2016: 2504-2505). Narrative binding, like representing global connectedness, is a necessary condition for representing causal relations in a bearing on manner. In order to represent that C* causes C in a bearing on manner, an agent must also represent that C* comes temporally before C and causes C.

In this vein, we can begin to respond to Trout's worry that, in the case of the proposed account of understanding, it doesn't rely on scientifically measurable psychological capacities. Representing in a bearing on manner depends on empirically measurable psychological capacities that can be disrupted with brain damage and that are underdeveloped in infancy.

To end this section, I compare my account with alternatives and respond to some objections. First, how might other accounts of grasping or understanding capture the relevant differences between the statements regarding Lewis's letter? I don't provide an exhaustive comparison. Rather, I discuss a couple alternative views. For Khalifa (2013), grasping involves "reliable explanatory evaluation." As Khalifa explains, "explanatory evaluator's inputs are various potential explanations of a phenomenon plus a body of relevant evidence, and their outputs are beliefs about which of these potential explanations is an actual explanation of this phenomenon" (2013: 6). Reliable evaluators are usually able to identify the correct explanation among competitors. This view of grasping does not help us delineate the intuitive difference between statements (1) – (4). The reason is that, given the interventionist criteria, (1) – (4) are all 'correct' explanations. Furthermore, (2) and (3), as I've argued, are plausibly as explanatorily 'deep' as the other. For these reasons, reliable explanatory evaluation cannot account for the difference.

Hills (2016) implicates the notion of "cognitive control" to account for grasping. For Hills, when one understands why P explains Q, one has the capacity to explain similar relationships, i.e., they have the relationship under their "control." If an explanatory relationship between P and Q is similar to P* and Q*, then provided with the information that P*, you might infer that Q* will follow. This kind of inferential capacity would not be present, according to Hills, if one merely

knows that P explains Q. This is related to how Ylikoski and Kuorikoski (2010, 2015) discuss understanding as an ability to make counterfactual inferences.[17]

By contrast to these accounts, my view of understanding is mentalistic. It is plausible that the mental state of understanding makes available the kind of inferential capacities discussed by Hills, Ylikoski, Kuorikoski and Levy. For example, we might say that "cognitive control," in Hills's sense, could be employed in (2) but not (1) or (3). As I discuss below, on my view, these capacities can be reliable public *indicators* of understanding. However, they should not be conflated with understanding itself. In this way, while the capacity to have "cognitive control" might indicate the relevant difference, the fundamental difference, the thing that on my view might engender such externally evaluable abilities, is the distinctive mental state that (2) makes available.

Lastly, I respond to a couple objections regarding a mentalistic account of understanding. Kuorikoski and Ylikoski say that "the criteria for attributing understanding are public. When judging whether someone understands something, people do not attempt to look into the person's mind; rather, they set out to observe whether he or she can make relevant counterfactual inferences about the phenomenon in question" (2015: 3820). While the folk attribution of 'understanding' to others depends on publicly available criteria (for an interesting discussion, see Wilkenfeldet al. 2016), this does not rule out that a theory of understanding can be mentalistic. Like other mental states, understanding plausibly makes new cognitive and behavioral capacities available to agents. But this doesn't mean that we should conflate these capacities with understanding itself. To do so is to endorse a behaviorist conception of understanding that I reject. Rather, public observance of the exercise of these capacities can function as a reliable indicator that subjects have activated a mental state constitutive of understanding.

Secondly, Kuorikoski and Ylikoski note that "scientific understanding is essentially collective. Scientific understanding proper is not what happens inside individual minds, but is constituted by the collective abilities of the scientific community to reason about and manipulate the objects of investigation" (2015: 3821). In response, my account promotes skepticism that scientific understanding can be collective in this way. Scientific understanding is constituted by the mental states of individuals. However, some might refer to the aggregation of these mental states as collective understanding. If collective understanding is construed in this way, then it is consistent with the proposed view. In addition, theorists who propose that there can be group

[17] For Levy (2020), understanding is possessing a representation that promotes such inferences.

mental states might also attribute understanding to the group. I don't take a stance on this possibility here. In any case, and most importantly on my view, the proposed account is consistent with the idea that scientific explanation, knowledge and progress are often collective endeavors.

Let us take stock. In this section, I've provided an account of scientific understanding. On this account, models that promote understanding allow agents to grasp how constituents of models bear on some explanandum phenomenon. Representing causal relations in a bearing on manner activates the same psychological kind of state as when we grasp how a story's conclusion is the product of events arranged in space and time.

So far, I've motivated the narrative intelligibility account by showing that it accounts for the intuitive difference between the statements involving Lewis's letter. In the next section, I further motivate this view by showing that it accounts for some cases in chemistry, psychology and biology. In these cases, some may wonder whether a *degreed* notion of understanding is more appropriate. On this plausible analysis, certain models might promote a low degree of understanding in comparison to a higher degree contrast model. On the narrative intelligibility view, a degreed notion of understanding could be cashed out as mental representations of causal relations involving more or less robust bearing on manners of representation. In these cases, the mental state contents might be described by: 'C* causes C in a bearing on manner to degree D.' While I am open to a degreed notion of understanding, I stick with binary phrasing below. In other words, I discuss models as if they either do, or do no, promote understanding for some agent. But some of these cases of a lack of understanding might be plausibly construed as cases of *thin* understanding. In such cases, the degree to which an agent represents that C* causes C in a bearing on manner can be considered low.[18]

4 Understanding and Mechanistic Explanations

The statements involving Lewis's letter are plausibly best conceptualized as *etiological models* which represent a set of counterfactual relationships among events.[19] How the distinction between explanation and understanding might surface in the context of *constitutive* explanations warrants further investigation. Constitutive models, as distinct from etiological models, tell us how some

[18] Even on a degreed notion of understanding, the discussion in Section 2 suggests that the degree of understanding that some model promotes would not be a function of the degree of explanatory power present in that model.

[19] Although, in Section 2, I discussed the possibility that the representations of these events could be construed as models of historical mechanisms.

phenomenon is produced by the organization of entities and their activities (Machamer et al. 2000, Craver 2007). Such models represent a mechanism, with its component parts, the parts' activities and the organization of the parts, where the phenomenon is the behavior or property of that mechanism. In what follows, I show how providing a representation of the mechanism for P is a powerful means toward understanding P. This is not to suggest that providing the mechanism for P necessarily entails understanding P. Nor is it to suggest that to understand P one must represent the mechanism for P. While much has been written about the importance of mechanistic explanation, the subsequent discussion illustrates that part of the reason they are important is because they often promote understanding.

To start, consider the ideal gas law: $pV = nRT$. When given a causal interpretation, the law represents counterfactual dependency relationships between pressure (p), volume (V), temperature (T), the amount of substance (n) and the ideal gas constant (R). Even in non-ideal circumstances, i.e., gases that are not in the limit of zero pressure, the equation can be used to represent relations among observable properties of actual gases. This is an example of an etiological model, which can be useful in determining, for example, why pressure and volume changed for some particular gas as a function of a rise in temperature.

Such an etiological model is explanatory given Woodward's interventionist framework. It answers a set of w-questions. However, it is a deficient source of understanding for somebody who does not grasp how a change in some gas's temperature causes an alteration in its volume. On the proposed account, this is because such a person cannot represent causal relations between variables in the law statement in a bearing on manner. This is indicated by the fact that, given the law statements by itself, we cannot comprehend how temperature increase *bears on* the ultimate increase in pressure. That psychological state is not made available to us given the information contained in the law statement alone.

By contrast, consider a constitutive model of the phenomenon that pressure is counterfactually dependent on temperature. This will involve representing a mechanism that models the interaction of gas molecules with each other and with the walls of the container, as well as the molecules' change in velocities. Once the mechanism of gas molecule interaction is provided, we can begin to grasp how rise in temperature alters volume and pressure. The mechanistic details fill in the narrative-like gaps necessary to promote understanding. Thus, this is

a case where understanding P, where P is the counterfactual dependence of volume on temperature for some gas, plausibly requires representing the mechanism for P.

Another illustrative example occurs in the context of Bayesian perceptual psychology. Broadly, we can think of the perceptual system as the capacity to represent distal features like shape, color, size, etc., on the basis of proximal sensory stimulation (e.g., simulation of light onto the retina). Bayesian modeling is a way to accurately describe the relationship between these sensory inputs and the perceptual representations that the system outputs. According to the Bayesian framework, our perceptual system draws an unconscious statistical inference. It provides perceptual estimates of distal properties in response to three factors. The first is *prior probability*. Based on experience, this is how often it is the case that these distal features are present in our environment independent of the current proximal evidence. Second, there is *prior likelihood*. This is how likely, given our past experiences, there is this distal feature given the current proximal stimulation. And lastly there is the *proximal stimulation* presented during some perceptual episode.

The Bayesian model of perception provides us with counterfactual dependency information about the explanandum phenomenon, i.e., the perceptual estimate. It tells us how the perceptual estimate would have been different given interventions to these three variables (e.g., if the subject were presented with different patterns of lighting directions in the past). Hence, under the interventionist criteria, Bayesian models are explanatory because they provide answers to a set of w-questions (see Rescorla 2015, 2018).

Like the ideal gas law, such Bayesian models are explanatory, yet they plausibly do not promote understanding. The Bayesian model on its own does not allow agents to grasp how past experiences *bear on* a perceiver's current representation of the distal environment. This requires looking at, for example, mechanisms of tuning in sensory cortices and predictions sent to the primary visual cortex (for a review, see de Lange et al. 2018). Understanding how prior experiences alter current distal representations plausibly requires learning about the neurobiological mechanisms that underpin the relationships represented in the Bayesian model.

Finally, consider a contrastive case between, not an etiological and constitutive model, but between two mechanistic models at different stages of development. Models of protein synthesis have progressed significantly from the 1950's. Craver and Darden (2013) compare James Watson's 1952 model of the protein synthesis mechanism, DNA→ RNA→ protein, with more complete models available today. According to them, Watson's model is an example of a

mechanism sketch. Mechanism sketches are incomplete descriptions of mechanisms. They have gaps in the description of components, or certain filler terms, or boxes and arrows that identify the parts of the mechanism that are not yet understood (Machamer et al., 2000, Craver 2007, Craver and Darden 2013). This leaves out important details regarding how the mechanism produces the phenomenon. As a mechanism sketch, it contains information about some parts of the mechanism, but it doesn't indicate how those parts interact. This omission prevents the mechanism sketch from promoting understanding. While it represents that protein synthesis is counterfactually dependent on DNA, we cannot mentally represent DNA as causing protein synthesis in a bearing on manner.

By contrast, a *mechanism schema*, according to Craver and Darden is, "a description of a mechanism, the entities, activities and organizational features of which are known with sufficient detail that the placeholding in the schema can be filled in as needed" (2013: 31). In the case of mechanism schemas, there are less explanatorily relevant components of the mechanism omitted in the model. This gets us closer to what Craver and Kaplan (2020) refer to as *Salmon-completeness*, i.e., where no explanatorily relevant component of the causal structure is left out of the model.

Contemporary models of protein synthesis represent more complete constitutive explanations. Such mechanistic models include more entities, e.g., nucleotides, ribosome, codons, etc., and more activities, e.g., unzipping, transcription, attaching, transfer, etc., which promote narrative intelligibility of protein synthesis in a way that Watson's earlier model did not. For example, a more complete model describes how the DNA unzips in the nucleus, mRNA nucleotides transcribe the DNA message, the mRNA goes to the ribosome attaches to it and a codon is read, etc. By providing a more complete mechanism, we can grasp how DNA and RNA bear on the eventual end-state of protein synthesis.

The upshot here is that, in the case of several scientific phenomena, a mechanistic model, or a more complete mechanistic model, is needed to achieve understanding of those phenomena. The above cases exhibit contrastive deficiencies psychologically analogous to those found in statements (1) – (4) as discussed above.

Experiments by Kriz and Hegarty (2003) and Hegarty et al. (2003) also elucidate the intelligibility that mechanistic models can provide. They investigated the role of "mental animations" in the intelligibility of machine behavior. This is the process by which motion is inferred, i.e., mentally visualized, on the basis of static (non-animated) diagrams. Evidence for

mental animations was found via the tracking of eye-movements, where subjects moved their eyes in the direction of the static diagram's direction of motion. Subjects demonstrated what the authors called "understanding" by answering certain troubleshooting questions about the machine. Correctly answering such questions demonstrated that subjects "understood" how the machine worked. Their studies showed that "understanding" was correlated with the use of mental animation. This is consistent with the proposed framework. To mentally animate a machine's causal behavior is to mentally visualize how certain parts of a machine, and their activities, bear on some end-state. Mental animation is thus one way in which causal explanatory relationships can be mentally represented in a bearing on manner. In this vein, mental animation can be construed as a species of narrative intelligibility.

How is it that mechanistic models are a powerful source of scientific understanding? I don't provide a comprehensive account. However, I'll make a few tentative suggestions. First, as mechanists such as Craver, Darden and Kaplan have discussed, mechanistic explanations don't only answer a set of w-questions, but *how-does-that-work* questions (or "h-questions") (Craver and Darden 2013, Craver and Kaplan 2020). It is plausible that being able to answer h-questions regarding P goes hand in hand with understanding P.

Second, mechanistic models often arrange component parts, and their activities, into a temporal order. This is also a feature of many etiological models. Temporal arrangements allow representing causal relations in a bearing on manner, where we can comprehend how the activities of certain parts contribute to and further along the explanandum phenomenon. Third, mechanisms describe *activities* that promote narrative intelligibility. Such activities can be construed as fundamental pieces of narrative glue that show *how* mechanistic parts causally interact with one another. For example, the activities of unzipping, transcription, attaching, etc., allow agents to represent causal relations between parts of the protein synthesis mechanism in a bearing on manner.[20] The representation of activities in mechanisms is crucial for grasping how components of the model hang together to further along some phenomenon.

[20] Some activities of mechanical parts might be construed as metaphors (see Levy 2020). Levy notes that it is common for models in biology, e.g., models of causal interactions between glands and muscles, to be metaphorically redescribed, e.g., as a gland "sending a message" to the muscle. Metaphors like these help agents grasp how parts of mechanisms interact with one another by relating the interaction to familiar interactions (like those found in narratives). Are such metaphorical activities part of the causal structure of the world? Or are such descriptions of activities merely a mental way of making real causal part interactions intelligible? I leave these complicated questions for future discussion.

Conclusion

In this paper, I've argued that explanation and understanding are distinct scientific achievements. As noted above, this contrasts with the opinion of many interventionists (Woodward 2003, Ylikoski and Kuorikoski 2010, Kuorikoski and Ylikoski 2015). I argued for this claim by showing that models can be explanatory yet fail to promote understanding. My positive account of scientific understanding implicates the notion of narrative intelligibility. According to the proposed view, when some model M promotes understanding for some agent A, the constituents of M can be mentally represented by A as causing the explanandum phenomenon P that M explains in a bearing on manner. When M promotes understanding, M makes a mental state available that is the same psychological kind as when we grasp events in a narrative as bearing on some ultimate conclusion. Finally, I suggested that mechanistic explanations, while not a necessary condition, are often a powerful source of scientific understanding.

References

Ash, S., Moore, P., Antani, S., McCawley, G., Work, M., & Grossman, M. (2006). Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology*, 66(9), 1405-1413.

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, 84(2), 214-233.

Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science.*

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. MIT press.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of science*, 68(1), 53-74.

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.* University of Chicago Press.

Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 199(3), 8807-8828.

Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, 71(1), 287-319.

Cummins, Robert. (1975). Functional Analysis. *Journal of Philosophy.* 72: 741-765.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in cognitive sciences*, 22(9), 764-779.

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

Dowe, P. (2000). *Physical Causation.* Cambridge University Press.

Elgin, C. (2007). Understanding and the facts. *Philosophical studies*, 132(1), 33-42.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, 71(1), 5-19.

Friedman, M. (1988), "Explanation and Scientific Understanding", in J. C. Pitt (ed.) *Theories of Explanation*, New York: Oxford University Press, 188–198.

Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis*, 72, 251-266.

Goudge, T. A. (1961). *The Ascent of Life.* Toronto: University of Toronto Press.

Grimm, S. R. (2010). The goal of explanation. *Studies in History and Philosophy of Science Part A*, 41(4), 337-344.

Grimm, S. R. (2014). Understanding as knowledge of causes. In *Virtue epistemology naturalized* (pp. 329-345). Springer, Cham.

Halford, Graeme S., William H. Wilson, and Steven Phillips. 1998. "Processing Capacity Defined by Relational Complexity: Implications for Comparative, Developmental and Cognitive Psychology." *Behavioral Brain Sciences* 21: 803–31.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and instruction*, 21(4), 209-249.

Hempel, Carl G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

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

Kriz, S., & Hegarty, M. (2003). Staring into space: Evaluating mechanical motion, mental animation and eye movements. Paper presented at the Twelth European Conference on Eye Movements, Dundee, UK.

Keven, N. (2016). Events, narratives and memory. *Synthese*, 193(8), 2497-2517.

Keven, N., Kurczek, J., Rosenbaum, R. S., & Craver, C. F. (2018). Narrative construction is intact in episodic amnesia. *Neuropsychologia*, 110, 104-112.

Khalifa, K. (2012). Inaugurating understanding or repackaging explanation? *Philosophy of Science*, 79(1), 15-37.

Khalifa, K. (2013). Understanding, grasping and luck. *Episteme*, 10(1), 1-17.

Khalifa, K. (2017). *Understanding, explanation, and scientific knowledge.* Cambridge University Press.

Kitcher, P. (1981). Explanatory unification. *Philosophy of science*, 48(4), 507-531.

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, 192(12), 3817-3837.

Kvanvig, J. L. (2003). *The value of knowledge and the pursuit of understanding.* Cambridge University Press.

Levy, A. (2020). Metaphor and Scientific Explanation. In *The Scientific Imagination*. Oxford University Press.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556-567.

Lewis, D. K. (1986a). Causal explanation. In *Philosophical Papers: Volume II.* Oxford: Oxford University Press.

Lewis D. K. (1986b) Postscript c to 'causation': (insensitive causation). In *Philosophical Papers: Volume II.* Oxford: Oxford University Press.

Lipton, P. (2009). Understanding without explanation. *Scientific understanding: Philosophical perspectives*, 43-63.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, 67(1), 1-25.

Mink, L. O. (1970). History and fiction as modes of comprehension. *New literary history*, 1(3), 541-558.

Morgan, M. S., & Wise, M. N. (2017). Narrative science and narrative knowing. Introduction to special issue on narrative science. *Studies in History and Philosophy of Science* Part A, 62, 1-5.

Pritchard, D. (2014). Knowledge and understanding. In *Virtue epistemology naturalized* (pp. 315-327). Springer, Cham.

Prychitko, E. (2021). The causal situationist account of constitutive relevance. *Synthese*, 198(2), 1829-1843.

Rescorla, M. (2015). Bayesian perceptual psychology. *The Oxford handbook of the philosophy of perception*, 694-716.

Rescorla, M. (2018). An interventionist approach to psychological explanation. *Synthese*, 195(5), 1909-1940.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science*, 26(5), 521-562.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Salmon, W. C. (1998). *Causality and explanation*. OUP.

Salmon, W. C. (1989). *Four decades of scientific explanation.* University of Minnesota Press.

Scriven, Michael. (1962). Explanations, predictions, and laws. *Minnesota studies in the philosophy of science*, (3), 170-230.

Siegel, G., & Craver, C. F. (2024). Phenomenological Laws and Mechanistic Explanations. *Philosophy of Science*, 91(1), 132-150.

Strevens, M. (2013). No understanding without explanation*. Studies in history and philosophy of science Part A*, 44(3), 510-515.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, 69(2), 212-233.

Trout, J. D. (2017). Understanding and Fluency. In *Making Sense of the World: New Essays on the Philosophy of Understanding* ed. Grimm S. R., 233-251.

Velleman, J. D. (2003). Narrative explanation. *The philosophical review*, 112(1), 1-25.

Wilkenfeld, D. A., Plunkett, D., & Lombrozo, T. (2016). Depth and deference: When and why we attribute understanding. *Philosophical Studies*, 173(2), 373-393.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford university press.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, 25(3), 287-318.

Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part II: Plumbing explanatory depth. *Noûs,* 37(2), 181-199.

Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical studies*, 148(2), 201-219.