Explanatory Norms and Interdisciplinary Research

Chiara Lisciandra

Draft: Please do not circulate or quote without permission

Abstract

This paper provides resources from the philosophy of science to identify differences between explanatory norms across disciplines and to examine their impact on interdisciplinary work. While the body of literature on explanatory norms is expanding rapidly, a consensus on a theoretical framework for systematically identifying norms across disciplines has yet to be reached. The aims of this paper are twofold: (i) to provide such a framework and use it to identify and compare explanatory norms across different domains; and (ii) to derive indications about interdisciplinary practice accordingly. By pursuing these goals, this work aims to be both theoretically significant and practically relevant. It contributes to the ongoing work on explanatory norms; and offers recommendations for the analysis of interdisciplinary science.

Keywords: Explanatory Norms; Dimensions of Explanatory Power; Interdisciplinary Science; Model Transfer; Integration.

1 Introduction

In recent decades, science has become increasingly interdisciplinary, as evidenced, for instance, by the proliferation of interdisciplinary programs and research centers worldwide. Notable examples include the *International Panel for Climate Change* (IPCC) and the *European Council for Nuclear Research* (CERN), where hundreds of scientists with different backgrounds, training, and expertise work together on joint problems (Frodeman et al. 2017).[1] A central motivation for interdisciplinary science is its potential to address some of the most significant challenges of our times—such as climate change, global health, and inequality—which are too broad and complex for individual disciplines to tackle in isolation (NSF 2020; Tuana 2013).[2]

In addition, university boards and funding agencies are calling for interdisciplinary science, and introducing new interdisciplinary educational and research programs and funding schemes. Here, interdisciplinary science proceeds at the level of fundamental and applied research to tackle complex scientific problems that ask for the concerted effort of scientists from different fields.[3]

The increasing academic and societal demand for interdisciplinarity, however, cannot be met fully without addressing a wide range of scientific and institutional challenges that it raises (Mäki 2016).[4]

One of the main obstacles to collaboration across disciplines is that scientific practices tend to be strongly domain-specific (Fagan 2019, MacLeod 2018). Scientists are generally trained within disciplinary boundaries that structure scientific work and regulate problem solving. Different domains even have distinct *explanatory norms*, in other words different criteria for what counts as adequate explanatory claims. Some

[1]The IPCC and the CERN are two distinct examples of scientific collaboration. The IPCC includes scientists from a wide range of fields spanning climatology, geophysics, and social sciences, while CERN's composition is more cohesive, consisting mostly of experimental and theoretical physics, but also astrophysics, engineering, and computer science (For more information on the interdisciplinary composition of one of the main groups at CERN, the Compact Muon Solenoid group, see the description of the group on the CERN website here).

[2]Against a characterization of interdisciplinarity as a means-end relationship, see Mäki (2016). According to him, interdisciplinary science should be defined in terms of "whatever relevant relationship [there is] between two or more scientific disciplines or their parts."(p. 331)

[3]For an analysis of some of the main factors motivating interdisciplinary work see, for example, Bechtel (1986) and Darden and Maull (1977). For a recent treatment, see the thorough discussion on the interdisciplinary relationship between evolutionary and developmental biology (Evo-Devo) by Love (2021).

[4]For an overview of institutional challenges to interdisciplinary work see, for example, Huttoniemi and Rafoìs (2018).

domains strive for generality and tractability, while others focus on specificity and accuracy. Since one of the main goals of science is to explain phenomena, if scientists disagree on something as fundamental as explanatory adequacy, it is likely that significant clashes will emerge in the context of interdisciplinary collaboration.

Explanatory commitments often become apparent through interdisciplinary practice, leading to a wide array of epistemological issues. For instance, there are questions about whether scientific domains with different explanatory norms can engage in interdisciplinary exchange, how disciplines that endorse different explanatory standards can contribute to an overall explanation, and how to reconcile criteria of explanatory adequacy when these pull in different directions (Brigandt 2013a, Fagan 2019, Herfeld and Lisciandra 2019, Love 2012).

This paper aims to shed light on the role of explanatory norms in interdisciplinary science by bringing together two distinct research areas that have so far mostly developed independently. On the one hand, the current literature on the methodology of interdisciplinary science identifies explanatory norms and examines their interaction in interdisciplinary practice. For example, some disciplines prioritize explanations based on the accuracy of collected data, while others favor explanations based on abstract mathematical models—and they may experience difficulties when working together. This literature acknowledges that explanatory standards play a crucial role in channeling and constraining interdisciplinary exchange (Fagan 2016) but also that similar combinations of explanatory norms lead to similar patterns of interdisciplinary interaction (MacLeod 2018). Additionally, this literature closely examines scientific practice and identifies explanatory standards in a manner that is sensitive to disciplinary specificities.

On the other hand, the general philosophy of science literature on explanation suggests that explanatory standards are, at least in some cases, instances of *general* dimensions of explanatory power—features that characterize the quality of explanation in the abstract (Ylikoski and Kuorikoski 2010). Examples include insensitivity, precision, and integration. While this approach acknowledges the plurality of explanatory norms in science, some of which are unique to specific domains, it also identifies certain *qualities* of explanation that typically pertain to particular domains or sets of domains. This literature suggests that there may be an overarching, albeit not exhaustive, set of explanatory norms, with certain domains usually endorsing specific elements of this set and others different elements. This literature thus promises to offer a theoretical framework that can help guide the analysis of explanatory norms across domains.

In this paper, I will argue that combining these two approaches allows for a better understanding of explanatory norms and their role in interdisciplinary science. By studying the explanatory norms of a discipline—identifying which ones it emphazises and which ones it de-emphazises—we obtain what we might think of as a discipline's profile of explanatory norms. Similar profiles facilitate interdisciplinary science, while differing ones may create obstacles to interdisciplinary activities. By deepening our understanding of explanatory norms, this paper aims to help interdisciplinary science address some of the challenges it faces due to differences in explanatory norms.

The paper is structured as follow: The next two sections (Sections 2 and 3) set the stage by providing an overview of recent philosophical literature on explanation and interdisciplinarity. Section 4 delves into the notion of explanatory norms, arguing that the literature would benefit from a theoretical account to study these norms across domains. In Section 5, I propose to draw on Ylikoski and Kuorikoski's (2010) account of dimensions of explanatory power for this purpose. I will argue that dimensions of explanatory power capture a relevant set of explanatory norms and, in Section 6, show that such norms provide indications about the dynamics of interdisciplinary projects. Finally, Section 7 concludes.

2 Background

In this section, I first introduce briefly the notion of explanatory norms. Next, I argue for the combination of the literature from the philosophy of the special sciences and the general philosophy of science, showing how this combination can be fruitful for studying interdisciplinarity. Finally, I outline the main reasons for focusing specifically on explanation, rather than on other features of scientific inquiry that are nevertheless also relevant to interdisciplinary work.

2.1 Explanation

In a pre-theoretic sense, explanations are typically answers to why-questions or why things happen, "where the "things" in question can be either particular events or something more general–e.g., regularities or repeatable patterns in nature." (Woodward and Ross 2021). In the philosophical literature, different models of (scientific) explanation have been developed to characterize what qualifies as an explanation. For example, according to the Deductive-Nomological (DN) model, explanations are logical arguments in which the explanandum is shown to follow logically from a law of nature, as well as from premises specifying initial conditions. According to counterfactual accounts, explanations show how the effect would change if its causes were changed (Woodward 2005, p.11).

Depending on the model of explanation one adopts, different criteria define what counts as an explanation and what makes one explanation better than another—in other words, what determines the explanatory power of an explanation (Schupbach 2011). For instance, within the DN model, one explanation is better than another, the more it reduces the surprise of the explanandum given the laws of nature (Schupbach and Sprenger 2011). In a counterfactual account, an explanation is better than another when it can answer a larger set of "what-if-things-had-been-different" questions (Ylikoski and Kuorikoski 2010).

Ylikoski and Kuorikoski (2010) have dissected several dimensions of explanatory power that derive their justification within the counterfactual account of explanation. Explanations that satisfy such dimensions are considered to be better than those that do not, or do so to a lesser degree. For example, scientists often aims for *precise* explanations. Precision refers to the sharpness with which the explanandum is defined with respect to its contrast class. The more precise the explanandum, the larger the set of answers to "what-if" questions that the related explanans provides (more on this in Section 5).

This paper takes Ylikoski and Kuorikoski's account (2010) as a case study with which to analyze the role that explanatory dimensions play in interdisciplinary practice. This account illustrates how we can distinguish some scientific domains by the set of explanatory dimensions they adopt. As I argue in later sections, I consider these dimensions to function as the explanatory norms of those domains. My main claim is that, when explanatory norms align, this can facilitate interdisciplinary interaction, whereas misaligned norms may hinder collaboration between fields.

While I illustrate this claim within the context of Ylikoski and Kuorikoski's explanatory framework, the main claim of this paper does not hinge on that specific account. If a competing account were considered to be more suitable for defining the criteria for explanation and explanatory norms, the central point regarding what to expect in case of the alignment or misalignment of norms would remain valid.

Furthermore, the overall point of this paper applies even if one endorses a pluralistic view on scientific explanation, where multiple explanatory accounts coexist. In this scenario, conflicts between fields could be attributed to their adherence to different explanatory frameworks. The core idea would then extend to a higher-level norm of explanation that characterizes these frameworks: fields subscribing to the same explanatory framework are more likely to succeed in interdisciplinary interactions compared to those with entirely different explanatory frameworks.

2.2 General philosophy of science and the philosophy of the special sciences

The role of explanatory norms has gained increasing recognition in the philosophical literature, with a growing body of work focusing on the explanatory features that characterize individual disciplines and their impact on the development of interdisciplinary projects.[5]

To date, much of the literature has relied on case studies from specific fields, which offer a growing collection of examples drawn from scientific practice. For instance, Fagan (2016) examines a case study of (failed) interdisciplinary research in systems biology. In particular, she argues in particular that the discord between theoretical biologists and experimental researchers "is rooted in divergent views of explanation" (p. 873).

While case studies offer rich and detailed evidence, one of their primary limitations is that extrapolating explanatory norms from them can be a demanding process. Identifying norms specific to fields requires considerable effort, even before comparisons across fields can be made. To advance research in this direction, this paper draws on the philosophy of science literature, which offers general accounts of explanation and, in particular, of dimensions of explanatory power (Ylikoski and Kuorikoski 2010).

The theoretical approach does not replace but rather complements a more fine-grained, bottom-up analysis based on case studies and local comparisons. The literature focusing on case studies gathers evidence about explanatory norms and their role in interdisciplinary research. Meanwhile, a theoretical framework provides conceptual tools to inform and develop case study analyses further, potentially leading to revi-

[5] For explanations in biology see, e.g., Ross 2022; for explanations in economics, see, e.g., Lehtinen and Kuorikoski 2007, Marchionni 2022; for explanations in interdisciplinary projects see, e.g., Brigandt 2010, Fagan 2016, Green and Andersen 2019, Love and Lugar 2013.

sions of the theoretical framework itself. In this way, these two research areas work in tandem and can mutually inform each other.[6]

2.3 Explanatory norms, institutional norms, and other norms of scientific inquiry

Although this paper focuses on explanatory norms, there exists a wide array of other types of norms at work in scientific communities, which significantly impact interdisciplinary projects. For instance, institutional norms play a critical role in either facilitating or hindering interdisciplinary collaboration (see, for example, Huttoniemi and Rafoìs 2018 and Lyall 2019). Examples include assessment norms that align incentives and reward criteria across domains. Additionally, norms related to disciplinary identities, cultures, and values can sometimes supersede issues related to explanatory norms (O'Rourke and Crowley 2013, Salmela and Mäki 2018).

Moreover, even when focusing on norms of scientific inquiry, it is important to consider more than just explanation. There are also norms regarding description, prediction, classification, and measurement. Relatedly, the idea that explanations provide answers to why-questions is typically central to distinguishing explanations from descriptions and predictions (although see, for example, Churchland 1989); however, different models of explanation articulate this distinction in various manners (Woodward 2005). Regardless of the specific characterizations, difficulties can arise among disciplines that pursue different aims, such as explanation versus description or prediction.

Although not all scientific fields assign equal importance to explanation, explanation remains a central scientific activity and is particularly relevant in the context of interdisciplinary research. Working to explain the same phenomenon or complementary aspects of it often motivates disciplines to engage in interdisciplinary work (Darden and Maull 1977). In this context, when scientists disagree on the appropriate approach for achieving explanatory goals, they may view the methods employed by other domains as irrelevant or misplaced. Differing conceptions of what constitutes explanation can therefore create difficulties in joint projects.

In light of this, this work focuses on explanatory norms, albeit recognizing that such

[6]Notwithstanding their complementarity, the variety of explanatory norms is such that not all of them may be subsumed under general dimensions. Therefore, while there is an overlap between the two approaches considered in this paper, each side may ultimately retains its own specificities.

norms likely interact with other kinds of norms—both scientific and institutional—in complex ways. By specifically examining explanatory factors, this paper aims to take an initial step toward a broader analysis of how explanatory norms intersect with other kinds of norms, while acknowledging that some effects of the overall set may only be captured in the aggregate.

3 Interdisciplinarity

The claim of this paper is that the alignment or divergence of explanatory norms between different domains can facilitate or hinder interdisciplinary work. This section briefly discusses the notion of interdisciplinarity and, relatedly, of disciplines and sub-disciplines as units of scientific inquiry.

While there are different views on interdisciplinarity, in this paper I consider it to exist on a spectrum. At one end of the range, an interdisciplinary exchange occurs when objects—such as models, theories, concepts, instruments, or methods—"move" from one domain to another, by importation or exportation, to address issues in one of the domains involved. At the other hand, interdisciplinary activities lead to the integration of different domains, or specific aspects of them, such as models or concepts, to tackle novel problems that require contributions from both sides.[7]

One of the reasons for this stepwise notion of interdisciplinary is that, even in the case of transfer, disciplines use resources from different fields to advance knowledge within their own domain.[8] Importation and exportation are crucial features distinguishing interdisciplinarity from other forms of interaction across domains, such as multidisciplinarity, where different disciplines pursue a common goal but remain separate throughout the process. Later in this section, I will argue that explanatory norms play a significant role even at the level of interdisciplinary transfer. By showing that explanatory norms are relevant even in this "thin" sense of interdisciplinarity, we have stronger reasons to believe that they also affect "thicker" forms of interdisciplinary interaction.

Any definition of interdisciplinarity crucially depends on what *disciplines* are, a

[7]For an overview on interdisciplinarity and integration, see the introduction to the Special Issue *Integration in Biology: Philosophical perspectives on the dynamics of interdisciplinarity* edited by Brigandt (2013b) and the contributed papers in it.

[8]See on this, Grüne-Yanoff and Mäki 2014. For case studies on interdisciplinarity without integration, see, e.g., Grüne-Yanoff 2011, and MacLeod and Nagatsu 2016.

topic central to a longstanding debate. In the literature, different authors use different terms to denote scientific clusters, such as areas, disciplines, domains, fields, paradigms, and specializations.[9] This literature considers how to partition the scientific landscape into units of analysis so that research that takes place across them qualifies as interdisciplinary (Bechtel 1986, Darden and Maull 1977, Kuhn 1970, Laudan 1977). While some authors refer to fields as sets encompassing research questions, explanatory factors, methods, and techniques (Darden and Maull 1977), others view disciplinary domains as sets of closely related cognitive resources (Andersen 2016, McLeod 2018). Still others emphasize their sociological and historical dimensions (Bechtel 1986, Fagan 2019) In this paper, I use the above terms interchangeably and remain neutral regarding the previous accounts. The main reason is that, in contemporary science, scientific units exist at various levels of aggregation, making it difficult to define clear boundaries between them. Although different criteria—scientific, cognitive, sociological, etc.—can be used to characterize specific scientific units, the same criteria might lead to different, intersecting distinctions depending on the level of granularity one uses.

An interdisciplinary project can thus refer to exchanges across traditionally conceived disciplines, such as physics, economics, and chemistry, or between sub-domains across disciplines, or even across domains within the same discipline, as in the case of experimental and theoretical biology, or micro- and macro-development economics. The main point is that such cases represent situations where scientific units that typically work separately, come together to work on joint projects. In such collaborations, they may face significant challenges in combining or reconciling different or opposing standards.

A recent focus in the philosophical literature on interdisciplinarity is *interdisciplinary model transfer*, which involves using the same models across scientific domains to address questions other than those for which these models were originally developed (Humphreys 2019). A standard example is the Lotka-Volterra model, which originated in population ecology and has then been applied in fields such as economics and medicine (Knuuttila and Loettgers 2016). Other examples include mathematical models from game theory and expected utility theory, the Ising model in physics, or the Barabási-Albert preferential attachment model (Herfeld and Dohne 2019, Knuuttila and Loettgers 2017, Nagatsu and Lisciandra 2021).

[9]Yet others include "systems of practice" (Chang 2012) and "repertoires" (Ankeny and Leonelli 2016).

Marchionni (2013) has shown that the application of formal models across domains is influenced by the specific *explanatory standards* characterizing each domain. In other words, the specification of theoretical machinery across domains is not solely dependent on their subject matter and the empirical content. The same modeling tools can be implemented differently in distinct domains, depending on the explanatory values that characterize those domains.

Marchionni (2013) illustrates this point through a case study from network theory. She examines the application of network theory respectively in economics, physics/applied mathematics, and analytic sociology. According to her analysis, the same theory has been implemented differently across fields, in ways that reflect their specific explanatory strategies. For instance, both economics and sociology are interested in network formation, i.e., the properties by which networks develop. However, economists explain network formation according to general principles of rational choice theory, while sociologists use explanations that refer to social norms and behavioral rules. Economists consider sociological explanations based on social norms to be ad-hoc, i.e., relying on very specific behavioral rules. Conversely, they favor explanations that are *general*—applicable to a broader set of phenomena—and *unified*—where the phenomena should derive from a limited set of axioms.

Marchionni's analysis is relevant in the context of interdisciplinary work as it identifies some limits in using a single formal template as a common basis for research across domains. While Marchionni emphasizes that different explanatory norms might interfere with this process, it is also possible that when there is at least some overlap between explanatory norms, this can facilitate their interaction.

To explore the the role of explanatory norms in interdisciplinary research, the next section focuses on the philosophical literature that analyzes the concept of explanatory norms; following that, it will move to introduce a framework on dimensions of explanatory power and show how this framework can be applied to interdisciplinary projects.

4 Explanatory norms

Explanatory norms are the *implicit* and *explicit* rules that govern what scientists consider to be explanatorily adequate within a specific domain. They determine the *goodness* of an explanation, in other words what makes an explanation a better answer to a why-question than other possible answers. Typical examples include accuracy, generality, tractability, and precision.[10] These norms identify the characteristics that a scientific explanation is expected to exhibit within a scientific field and guide actions aimed at fulfilling them accordingly. They are explicit when stated, for instance, in textbooks and research guidelines, while they are implicit when they are conveyed by practice and constitute a part of a scientist's background of tacit knowledge.

The notion of explanatory norms has been discussed in recent philosophy of science literature. According to Longino (1990): "Scientific practice is governed by *norms* and values generated from an understanding of the goals of scientific inquiry. [They put constraints on] *what counts as a good explanation*, for example, the satisfaction of such criteria as truth, accuracy, simplicity, predictability, and breadth" (p. 4, italics added).

Longino's work builds on Kuhn's well-known list (1977) of epistemic values. According to Kuhn, these values are the "standard criteria for evaluating the adequacy of a theory" (p.322). For instance, he argues that a theory should have broad *scope*, meaning it should be able to explain more facts and observations than initially intended. *Fruitfulness* refers to the theory's ability to generate new research findings and "disclose new phenomena or previously unnoticed relationships" (p. 322).[11]

While Kuhn uses the term *values*, in this paper, I adopt the notion of explanatory norms to refer to the qualities of an explanation. One of the main reasons for choosing the term norms is that, unlike values and virtues, which typically pertain to individual subjects, norms denote group-level features that characterize scientific communities. And similar to norms in social groups, explanatory norms become particularly visible when scientific groups with different norms interact with each other.

Moreover, I argue that explanatory norms qualify as *norms* in two respects: i) they are shared among (the majority) members of a scientific community; and, ii), they underwrite normative judgments, meaning explanations that adhere to these norms are considered superior to those that do not.[12]

Concerning the first condition, the literature on explanation reveals that differ-

[10] For more examples, see Kuhn's well-known list (1977) of epistemic values, which includes *accuracy, scope, simplicity, consistency, and fruitfulness.*

[11] As Kuhn observes, the list is not exhaustive. Moreover, Kuhn notes that each item is rather imprecise and can be interpreted differently. Finally, different items can conflict with one another: for instance accuracy and scope are typically competitive values in the evaluation of a scientific theory.

[12] The two conditions can be mapped to those used to identify social norms in social philosophy accounts (Bicchieri 2005).

ent scientific domains prioritize specific sets of explanatory norms over others[13]. For instance, in many subfields of economics, *tractability*—defined as the formulation of models that are analytically tractable—is a central requirement. However, in other domains, such as certain subfields of psychology, including clinical psychology, tractability does not play the same crucial role. It is important to note that tractability qualifies as an explanatory norm because it pertains to the analytical derivation of an effect from a set of assumptions or foundational principles. This is of value in a number of explanatory frameworks, for instance because the result is in agreement with the overall background theory. However, "allegiance" to explanatory norms is not rigidly fixed: within the same field, different explanatory standards may be adopted at different times or for different purposes (Brigandt 2013a). What matters is the consistency of scientists' beliefs and expectations regarding the "appropriate" norms in a given context, which identifies the norms in place.

Concerning (ii), the normativity of explanatory norms derives from their explanatory power. The basic idea is that explanatory norms retain normative force depending on whether they track explanatory power. In other words, adherence to explanatory norms leads to an increase in the explanatory power of an explanation, and whether something qualifies as an explanatory norm is justified within the account of explanation one defends. In what follows, I will explore the dimensions proposed by Ylikoski and Kuorikoski within the counterfactual account of explanation. Within this, each dimension is defended on the basis that it satisfies the main criteria that define an explanation in that account.

In this respect, the discussion of explanatory norms intertwines normative and descriptive philosophy of science. Ylikoski and Kuorikoski's framework serves as a benchmark model in this context.[14] It can be used as a normative model, assessing whether something qualifies as an explanatory norm and why; at the same time, it can be used as a descriptive model, insofar as explanatory norms aligns with scientific practice. Moreover, the model helps differentiate between norms that are not essentially related to explanatory power: some contribute to scientific progress through social or institutional processes or by promoting goals beyond scientific explanations; yet others may not appear to contribute to scientific progress but exist for path-dependence or

[13]See, e.g., Brigandt 2013a, Brigandt and Love 2012, Kuorikoski and Marchionni 2014, Lisciandra 2018, Love 2015, Love and Lugar 2013, MacLeod 2018.

[14]My thanks to an anonymous referee for suggesting to treat the account as a benchmark model that can be interpreted normatively and descriptively.

some other reasons. Finally, the benchmark model can also be challenged on the basis of the justification of its normative requirements. The following section illustrates how this approach can identify and compare explanatory norms across different scientific fields.

5 Dimensions of explanatory power

Ylikoski and Kuorikoski identify five dimensions of explanatory power: non-sensitivity, precision, factual accuracy, degree of integration, and cognitive salience. To illustrate, *non-sensitivity* indicates the degree to which an explanation is insensitive to changes in background conditions. The idea is that the less sensitive an explanation, the more powerful it is. *Factual accuracy* pertains to the idealizations that are considered to be adequate for an explanation, where the fewer the idealizations, the more powerful an explanation is (more below).

This account is embedded in the contrastive-counterfactual theory of explanation (Woodward 2005). This has found application in a wide-range of natural and "non-natural" sciences, including mathematics (Baron et al. 2020), economics, history (Brien 2013), and metaphysics (Schaffer 2016).

Within this framework, the dimensions serve as metrics to assess explanatory power, as explanations satisfying them provide answers to a broader range of *what-if-things-had-been-different* questions. In principle, this approach allows for comparisons between explanations either along a single dimension, such as precision; or across different dimensions, such as insensitivity versus accuracy. Either way, dimensions of explanatory power represent explanatory norms that influence interdisciplinary work, as I argue in what follows. To this aim, initially I examine which individual dimensions correspond to explanatory norms in particular domains, and then explore general attributes of the framework that make it suitable for analyzing the dynamics of interdisciplinary research.

At first glance, most of the dimensions identified in this account appear to be promising candidates for explanatory norms that could be linked to specific domains, or could at least be implemented in specific ways within those domains.[15]

[15] For the dimensions discussed, I refer the reader to Ylikoski and Kuorikoski (2010) for an analysis of how each one aligns with explanatory power, in other words how it produces explanations superior to those that either fail to satisfy the dimension or do so only partially. The purpose of this paper is not to critically engage with the authors' account, but to extend it by highlighting domain-specific

- Regarding insensitivity/precision, there tends to be a division between domains that prioritize one over the other. An explanatory relationship is considered less sensitive if it is more invariant to changes in background conditions or to interventions on the explanatory dependency. This dimension is particularly valuable in domains focused on high levels of abstraction, which therefore use formal methods to this end. It is thus a typical feature of disciplines that are interested in aggregate phenomena, as for instance those areas within economics and theoretical biology that develop equilibrium models and utilize mathematical modeling techniques. It should be noted that when epistemic goals align across domains, the level of insensitivity in their explanations can be compared. However, if these goals differ, different domains may refer to distinct sets of background conditions. As with other dimensions listed below, even when two fields subscribe to the same norm, they may satisfy it differently.

- Precision is a characteristic in domains that favor detailed characterizations of the *explanandum*. According to Ylikoski and Kuorikoski, it refers to the "sharpness" of the explanandum, i.e., the distinctiveness with which the explanandum is defined (p. 210). For example, an *explanans* that addresses a more specific explanandum is considered to be better than one addressing a less specific one. This is common in disciplines that focus on precise, specific phenomena, rather than broader explanations. Examples include analytical sociology, where researchers focus on the behavioral rules that generate specific phenomena within defined contexts and historical periods, and chemistry and molecular biology, where scientists aim to synthesize specific molecules. While it is true that changing the explanandum means no longer comparing the same phenomenon, the key point is that domains endorsing this norm may have specific views on what counts as a good explanandum and typically do not regard broader explananda as sufficiently explanatory.

- Factual accuracy concerns the idealizations included in an explanation, where fewer idealizations indicate a better explanation, assuming the level of abstraction remains constant. This dimension typically characterizes domains that aim to use tractable models. However, idealizations are ubiquitous across scientific domains, albeit with variations in their nature. Ylikoski and Kuorikoski give the example

features of relevance in the context of interdisciplinary exchange.

of Newtonian mechanics, which omits friction and air resistance when describing motion (p. 218). Similarly, models in population ecology and segregation models in social science follow this pattern. Even experimental and empirical sciences, which are more directly concerned with the details of the phenomena or the processes under study employ idealizations, albeit to refer to other aspects of their subject matter, such as assumptions underlying experimental procedures and the data analysis. Thus, while this dimension may manifest differently across contexts, certain disciplines may be more aligned in how they adhere to this norm compared to others. Nonetheless, the overarching idea remains that all fields embracing this norm are more closely related to each other than to those adhering to entirely different norms.

- Finally, the degree of integration measures how well an explanation connects with the broader body of knowledge within a certain discipline. In principle, an explanation that achieves a high degree of integration can reveal system-wide properties that a more isolated piece of evidence might not capture. This level of integration is typically achieved by disciplines that take a top-down approach, starting from a set of theoretical principles that guide their analysis, rather than a bottom-up approach seen in data-driven fields. Many disciplines find themselves at two opposite sides of a spectrum with respect to their degree of integration. For example, it is widely agreed that economics often begins with foundational principles and develops downwards, contrasting with psychology's more cumulative approach. A similar dichotomy exists in evolutionary systems biology, where evolutionary biologists focus on establishing general evolutionary principles while systems biologists concentrate on mechanistic explanations of specific traits (Green et al. 2015).[16]

The preceding analysis indicates how dimensions of explanatory power can serve as a framework for identifying explanatory norms, which in turn highlight similarities and differences between domains. As outlined, certain fields tend to adhere to specific norms more than others, although these norms may manifest differently across do-

[16] The list omits *cognitive salience* because what is deemed salient can be highly context-specific and challenging to analyze across domains. However, it is possible that certain disciplines share similar standards of salience, whatever they may be, making them more conductive to interdisciplinary exchange than those that do not share such standards. An example illustrating how a lack of cognitive salience impedes knowledge transfer can be found in Dais 2019.

mains. However, some degree of overlap can potentially lessen the divergence between disciplines regarding explanatory norms.

The explanatory profile of a discipline is not necessarily unique; two disciplines can share the same disciplinary profile, have some norms in common or even none at all. It is conceivable for two disciplines to adopt identical or similar explanatory norms while differing in subject matters, research questions, or other characteristics. For instance, consider mathematical physics, mathematical economics, and mathematics proper. Despite using similar methods and arguably sharing the same explanatory profiles, these disciplines are clearly distinguished by their intended subject matter.

It should also be noted that whether the list of dimensions is exhaustive or requires further amendment is a question that merits exploration in future research. The primary aim here is to assess how well the account serves as foundational framework for mapping domain-specific explanatory norms and their implementation. With ongoing refinements, the ultimate goal is to develop a blueprint to identify disciplinary profiles that encompass different sets of norms and their respective degrees.

In addition to its promise for studying explanatory norms, Yilkoski and Kuorikoski's framework is also well-suited for identifying features of interdisciplinary research and understand situations where different fields adopt different explanatory norms.

To begin with, their account provides a method to compare explanations of the same phenomenon based on different explanatory desiderata, i.e. while leaving the target system fixed. This situation frequently arises in interdisciplinary science when researchers approach a common scientific problem from their respective disciplinary perspectives.

Secondly, the framework does not address whether something explains in the first instance, as this relies on other criteria, particularly an interventionist account of causation. Instead, it focuses solely on identifying the dimensions that an explanation satisfies and, even if only informally, the degree to which it meets them. One explanation of a given phenomenon might just satisfy one dimension or a set of them, while another could satisfy a different set altogether. This flexibility allows different domains to provide explanations that vary in approach. In interdisciplinary projects this accommodates situations where different domains adhere to their own explanatory standards, assigning different weights to different (sets of) dimensions.

Thirdly, as the authors make clear, it is unlikely within this account that a single explanation satisfies all dimensions simultaneously, as some of them pull in opposite directions.[17] For instance, non-sensitivity typically involves a trade-off with precision, since a narrower explanandum increases sensitivity to interventions or changes in background conditions. Additionally, explanations with greater levels of integration may trade off factually accuracy and/or precision, as they require a higher degree of abstraction to reveal connections with other explanations. These trade-offs reflect a misalignment across domains, leading to tensions among scientists who uphold different explanatory standards. For instance, domains employing formal models are typically more open to (certain kinds of) idealizations than those relying on laboratory experiments (as discussed in the next section).

Finally, the account provides a framework within which to analyze cases of (scientific) disagreement. It enables the examination of disagreements arising from the adoption of different explanatory norms, as well as the distinction of those deriving from different kinds of scientific or institutional norms (Ylikoski and Kuorikoski 2010, p. 218).

In conclusion, the preceding analysis illustrates how the study of dimensions of explanatory power can serve as a framework for identifying explanatory norms, and to recognize variations of explanatory norms as differences in dimensions of explanatory power and their instantiations.

6 From explanatory norms to interdisciplinary research

The aim of this section is to show how the previous analysis of explanatory norms offers indications of patterns of interdisciplinary interaction. The idea is that alignment in explanatory norms, or their practical implementation, can foster collaboration across domains. Conversely, significant divergence in norms may hinder it.

To clarify, facilitating or hindering interdisciplinary interaction does not imply that the alignment or collision of explanatory norms can predict whether the ultimate success or failure of such interaction in terms of scientific progress. Rather, the point is that, provided that a certain interdisciplinary activity is considered to be promising in terms scientific gains, certain combinations of explanatory norms may allow scientists to collaborate more effectively compared to other combinations requiring a greater

[17] This point on the was already raised by Kuhn in his discussion on epistemic values.

overall investment of time and resources.

Secondly, and related to the previous point, this paper does not analyze which combinations of explanatory norms are more conducive to scientific advancement. It is possible that to generate interesting results explanatory norms need to be at an optimum distance from each other, neither too close nor too far apart (see, e.g., Love 2021). However, this issue goes beyond the scope of this discussion, which focuses instead on anticipating possible clashes or favorable matches among scientists based on their explanatory norms.

With this in mind, I will briefly present two examples where explanatory norms influence interdisciplinary research: the first illustrates a case where explanatory norms diverge, and the second where they partially align. The cases are illustrated both through the scientists' own methodological reflections and through an analysis of their scientific practice.

1. Micro/macro-development economics. To begin with, the literature has so far focused extensively on cases where different explanatory norms or frameworks tend to hinder interdisciplinary work (Fagan 2016, Fam and O'Rourke 2020). An example is development economics. It is well-documented that this field has long been divided between two main subdomains: macro-development economics and micro-development economics (Rodrik 2009). Both areas address development issues, such as reducing poverty and enhancing growth and living standards, but their approaches differ substantially. Macro-development economics examines the role of structural transformations, fiscal macro-policies, and international trade; microeconomics is mainly concerned with health, education, and fiscal micro-policies.

The misalignment between these two subfields can largely be attributed to differences in explanatory norms. Micro-development economists defend the *precision* achieved by small-scale interventions based on randomized-control trials (RCTs) (Rodrik 2009, p. 26). In contrast, macro-development economists criticize RCTs for their lack of generality (p. 26) and their *low degree of integration* with an overarching theory. Indeed, while macro-development economics typically follows a top-down approach, starting from macroeconomic theory and moving to observations, micro-development economics adopts a bottom-up approach, collecting evidence primarily through RCTs. The core of the debate is not merely the use of cross-country regressions in macro-development economics and of RCTs in micro-development economics, but rather the *validity of the explanations* afforded by these distinct methods. Despite calls for greater collaboration from scholars in both fields, reconciling these approaches remains challenging.

2. Gravity model in economics. The second example concerns the exchange of models from physics to economics, specifically the development of so-called *gravity equations* in international trade theory. The gravity model is just one of several instances where models have "migrated" from physics to economics. Other notable examples include the Ising model, which has moved from physics to the social sciences, and the kinetic models from statistical mechanics, which have been used to study the distribution of wealth (Bradley and Thebault 2019).

The central concept underlying the gravity model in economics is to model international trade flows by analogy with Newtonian gravity: the trade flow between two countries is proportional to the product of their "economic masses" and inversely proportional to their distance. This idea gives rise to the simple gravity equation:

$$X_{i,j} = g\frac{X_i X_j}{d_{i,j}}, \tag{1}$$

where $X_{i,j}$ is the bilateral trade flow between countries i and j, g is a "gravitational constant" of proportionality, X_i and X_j are the "economic masses" of i and j, and $d_{i,j}$ is the "distance" between the two countries. These variables can be instantiated in various ways, but in more simple gravity models, X_i, X_j are simply the GDP's of i and j, and $d_{i,j}$ is their physical distance.

Gravity equations were introduced into trade theory as an econometric model by Tinbergen (1962) and are considered among the most successful empirical models in economics (Anderson 2011, p.13). Despite their empirical success, many economists were initially skeptical of gravity models because of their perceived lack of a solid theoretical (microeconomic) foundation. As Leamer and Levinsohn (1995, p. 1387) noted, "[T]he gravity models are strictly descriptive. They lack a theoretical underpinning so that once the facts are out, it is not clear what to make of them."

In response to this, considerable effort has been devoted to providing gravity equations with theoretical foundations in such a way as to make them compatible with the background of theoretical knowledge in international trade theory. Several economists have worked to prove how gravity equations can be derived from particular models of international trade—such as a Ricardian model (Eaton and Kortum 2002), Heckscher-Ohlin models (Deardorff 1998), and from models of comparative advantage (Anderson and Van Wincoop 2003).

The effort that economists dedicated to providing theoretical foundations for gravity equations can be understood by considering the importance that economists place on explanatory norms such as *integration* with the overall body of knowledge.

7 Conclusions

This paper has focused on a particular aspect affecting the way that disciplines interact with each other: explanatory norms. The reason for this focus is that, first, in the last decades, the philosophy of the special sciences literature has provided increasing evidence on the explanatory features that characterize specific fields. Furthermore, this growing body of work shows that explanatory features cut across domains—their research questions, models, specific methods, and more. (Andersen 2015).

This paper aims to deepen our understanding of explanatory standards by synthesizing evidence on explanatory norms from previous local studies, and suggesting that they can be brought together in a unified framework. It proposes to apply Ylikoski and Kuorikoski's (2010) framework on dimensions of explanatory power to this goal. In their model, explanatory dimensions serve as abstract qualities that characterize explanations, including attributes like insensitivity, precision, integration, and others.

The claim of this paper is that specific combinations of these dimensions form a domain's explanatory profile. It argues that disciplinary profiles can share overlapping norms, and that closer alignment in these profiles facilitates interdisciplinary interaction compared to when they are more distant. Finally, since these dimensions derive from a theory of explanation, it becomes clear how different domains can apply different criteria for justifications, which are not inherently inconsistent, despite pulling in different directions.

Since dimensions of explanatory power are general features that abstract away from specific details characterizing explanations in particular fields, they illuminate what explanations have in common across disciplines. However, the analysis operates at a more general level than the in-depth examinations of specific domains found in the philosophy of the special sciences. Moreover, the list of explanatory norms presented may require further expansion and revision in light of ongoing theoretical work and additional case studies.

Finally, it is hoped that this work has provided some initial guidance on assessing interdisciplinary work by leveraging explanatory distances across domains. For example, in projects yet to be undertaken, it provides a means to anticipate potential difficulties that may arise in interdisciplinary collaborations and consider strategies to address them. For completed projects evaluated ex-post, it offers a way to acknowledge the challenges encountered due to explanatory differences.

In conclusion, this work shows that the norms that set the adequacy of an explanation the adequacy of an explanation vary across domains and influence collaborative projects accordingly. In light of the interdisciplinary turn in contemporary science, philosophers of science are uniquely positioned to contribute to this research. Their expertise in methodological issues can provide valuable insights and guidance on scientific inquiry at the intersection of different fields.

References

Andersen, Hanne (2016). "Collaboration, interdisciplinarity, and the epistemology of contemporary science". In: *Studies in History and Philosophy of Science Part A* 56, pp. 1–10.

Anderson, James (2011). "The gravity model". In: *Annu. Rev. Econ.* 3.1, pp. 133–160.

Anderson, James and Eric Van Wincoop (2003). "Gravity with gravitas: A solution to the border puzzle". In: *American economic review* 93.1, pp. 170–192.

Baron, Sam, Mark Colyvan, and David Ripley (2020). "A counterfactual approach to explanation in mathematics". In: *Philosophia Mathematica* 28.1, pp. 1–34.

Bechtel, William (1986). *Integrating scientific disciplines: Case studies from the life sciences*. Vol. 2. Springer Science & Business Media.

Bicchieri, Cristina (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bradley, Seamus and Karim Thébault (2019). "Models on the move: Migration and imperialism". In: *Studies in History and Philosophy of Science Part A* 77, pp. 81–92.

Brien, James (2013). "The role of causation in history". In: *History in the Making* 2.1, pp. 72–81.

Brigandt, Ingo (2010). "Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology". In: *Erkenntnis* 73.3, pp. 295–311.

— (2013a). "Explanation in biology: Reduction, pluralism, and explanatory aims". In: *Science & Education* 22.1, pp. 69–91.

— (2013b). "Systems biology and the integration of mechanistic explanation and mathematical explanation". In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44.4, pp. 477–492.

Brigandt, Ingo and Alan Love (2012). "Conceptualizing evolutionary novelty: moving beyond definitional debates". In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 318.6, pp. 417–427.

Dais, Photis (2019). "The double transfer of thermodynamics: from physics to chemistry and from Europe to America". In: *Studies in History and Philosophy of Science Part A* 77, pp. 54–63.

Darden, Lindley and Nancy Maull (1977). "Interfield theories". In: *Philosophy of science* 44.1, pp. 43–64.

Deardorff, Alan (1998). "Determinants of bilateral trade: does gravity work in a neoclassical world?" In: *The regionalization of the world economy*. Ed. by Jeffrey A. Frankel. University of Chicago Press, pp. 7–32.

Eaton, Jonathan and Samuel Kortum (2002). "Technology, geography, and trade". In: *Econometrica* 70.5, pp. 1741–1779.

Fagan, Melinda Bonnie (2016). "Stem cells and systems models: Clashing views of explanation". In: *Synthese* 193.3, pp. 873–907.

— (2019). "Explanation, Interdisciplinarity, and Perspectives". In: *Understanding Perspectivism*. Routledge, pp. 28–48.

Fam, Dena and Michael O'Rourke (2020). *Interdisciplinary and transdisciplinary failures: Lessons learned from cautionary tales*. Routledge.

Frodeman, Robert, Julie Thompson Klein, and Roberto Carlos Dos Santos Pacheco (2017). *The Oxford handbook of interdisciplinarity*. Oxford University Press.

Green, Sara and Hanne Andersen (2019). "Systems science and the art of interdisciplinary integration". In: *Systems Research and Behavioral Science* 36.5, pp. 727–743.

Green, Sara, Melinda Fagan, and Johannes Jaeger (2015). "Explanatory integration challenges in evolutionary systems biology". In: *Biological Theory* 10.1, pp. 18–35.

Grüne-Yanoff, Till (2011). "Models as products of interdisciplinary exchange: Evidence from evolutionary game theory". In: *Studies in History and Philosophy of Science Part A* 42.2, pp. 386–397.

Grüne-Yanoff, Till and Uskali Mäki (2014). "Introduction: Interdisciplinary model exchanges". In: *Studies in History and Philosophy of Science Part A* 48.

Herfeld, Catherine and Malte Doehne (2019). "The diffusion of scientific innovations: A role typology". In: *Studies in History and Philosophy of Science Part A* 77, pp. 64–80.

Herfeld, Catherine and Chiara Lisciandra (2019). "Knowledge transfer and its contexts". In: *Studies in History and Philosophy of Science Part A* 77, pp. 1–10.

Humphreys, Paul (2019). "Knowledge transfer across scientific disciplines". In: *Studies in History and Philosophy of Science Part A* 77, pp. 112–119.

Huttoniemi, Katri and Ismael Rafoìs (2018). "Interdisciplinarity in research evaluation". In: *The Oxford handbook of interdisciplinarity*. Ed. by Robert Frodeman, Julie Thompson Klein, and Roberto Carlos Dos Santos Pacheco. Harvard University Press.

Knuuttila, Tarja and Andrea Loettgers (2016). "Model templates within and between disciplines: from magnets to gases–and socio-economic systems". In: *European Journal for Philosophy of Science* 6.3, pp. 377–400.

— (2017). "Modelling as indirect representation? The Lotka–Volterra model revisited". In: *The British Journal for the Philosophy of Science*.

Kuhn, Thomas (1970). *The structure of scientific revolutions*. Chicago University of Chicago Press.

— (1977). *The essential tension: tradition and innovation in scientific research*. Chicago: University of Chicago Press.

Kuorikoski, Jaakko and Caterina Marchionni (2014). "Unification and mechanistic detail as drivers of model construction: Models of networks in economics and sociology". In: *Studies in History and Philosophy of Science Part A* 48, pp. 97–104.

Laudan, Larry (1977). *Progress and its problems: Towards a theory of scientific growth*. Vol. 282. Univ of California Press.

Leamer, Edward and James Levinsohn (1995). "International trade theory: the evidence". In: *Handbook of international economics* 3, pp. 1339–1394.

Lehtinen, Aki and Jaakko Kuorikoski (2007). "Computing the perfect model: Why do economists shun simulation?" In: *Philosophy of Science* 74.3, pp. 304–329.

Lisciandra, Chiara (2018). "The role of psychology in behavioral economics: The case of social preferences". In: *Studies in History and Philosophy of Science Part A* 72, pp. 11–21.

Longino, Helen E (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton university press.

Love, Alan (2015). "Collaborative explanation, explanatory roles, and scientific explaining in practice". In: *Studies in History and Philosophy of Science Part A* 52, pp. 88–94.

— (2021). "Interdisciplinarity in evo-devo". In: *Evolutionary Developmental Biology: A Reference Guide*, pp. 407–423.

Love, Alan and Gary Lugar (2013). "Dimensions of integration in interdisciplinary explanations of the origin of evolutionary novelty". In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44.4, pp. 537–550.

Lyall, Catherine (2019). *Being an interdisciplinary academic: How institutions shape university careers*. Springer.

MacLeod, Miles (2018). "What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice". In: *Synthese* 195.2, pp. 697–720.

MacLeod, Miles and Michiru Nagatsu (2016). "Model coupling in resource economics: conditions for effective interdisciplinary collaboration". In: *Philosophy of science* 83.3, pp. 412–433.

Mäki, Uskali (2016). "Philosophy of interdisciplinarity. What? Why? How?" In: *European Journal for Philosophy of Science* 6.3, pp. 327–342.

Marchionni, Caterina (2013). "Playing with networks: how economists explain". In: *European Journal for Philosophy of Science* 3.3, pp. 331–352.
— (2022). "3 Social Aspects of Economics Modelling". In: *Methodology and History of Economics: Reflections with and without Rules*.
Nagatsu, Michiru and Chiara Lisciandra (2021). "Why Is behavioral game theory a game for economists? The concept of beliefs in equilibrium". In: *A Genealogy of Self-Interest in Economics*, pp. 289–308.
National Science Foundation (2020). *Interdisciplinary research*. NSF. URL: https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/.
O'Rourke, Michael and Stephen Crowley (2013). "Philosophical intervention and cross-disciplinary science: the story of the Toolbox Project". In: *Synthese* 190.11, pp. 1937–1954.
Rodrik, Dani (2009). "The New Development Economics: We Shall Experiment, but How Shall We Learn?" In: *What Works in Development?: Thinking Big and Thinking Small*. Ed. by Jessica Cohen and William Easterly. Brookings Institution Press, pp. 24–47.
Ross, Lauren N (2022). "Cascade versus mechanism: The diversity of causal structure in science". In:
Salmela, Mikko Erkki Matias and Ismo Uskali Mäki (2018). "Disciplinary emotions in imperialistic interdisciplinarity". In: *Scientific Imperialism*. Ed. by Uskali Mäki, Adrian Walsh, and Manuela Fernández Pinto. Routledge Studies in Science, Technology and Society. Routledge, pp. 31–50.
Schaffer, Jonathan (2016). "Grounding in the image of causation". In: *Philosophical studies* 173.1, pp. 49–100.
Schupbach, Jonah (2011). "Studies in the Logic of Explanatory Power". Available at http://d-scholarship.pitt.edu/id/eprint/7885. PhD thesis. University of Pittsburgh.
Schupbach, Jonah and Jan Sprenger (2011). "The logic of explanatory power". In: *Philosophy of Science* 78.1, pp. 105–127.
Tinbergen, Jan (1962). *Shaping the World Economy*. New York, NY: The Twentieth Century Fund.
Tuana, Nancy (2013). "Embedding philosophers in the practices of science: bringing humanities to the sciences". In: *Synthese* 190.11, pp. 1955–1973.
Woodward, James (2005). *Making things happen: A theory of causal explanation*. Oxford university press.
Woodward, James and Lauren Ross (2021). "Scientific Explanation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2021. Metaphysics Research Lab, Stanford University.
Ylikoski, Petri and Jaakko Kuorikoski (2010). "Dissecting explanatory power". In: *Philosophical studies* 148.2, pp. 201–219.