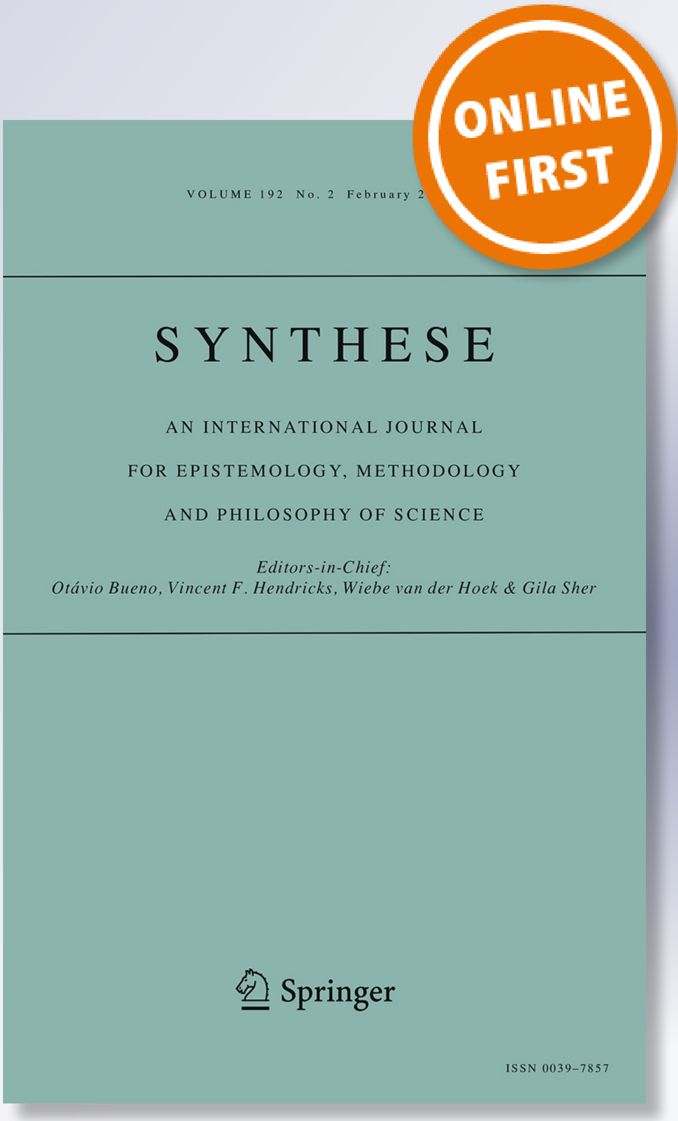
Goal-dependence in (scientific) ontology

David Danks

Synthese
An International Journal for Epistemology, Methodology and Philosophy of Science

ISSN 0039-7857

Synthese
DOI 10.1007/s11229-014-0649-1



Springer

Synthese
DOI 10.1007/s11229-014-0649-1

Goal-dependence in (scientific) ontology

David Danks

Received: 4 January 2014 / Accepted: 24 December 2014

Abstract Our best sciences are frequently held to be one way, perhaps the optimal way, to learn about the world's higher-level ontology and structure. I first argue that which scientific theory is "best" depends in part on our goals or purposes. As a result, it is theoretically possible to have two scientific theories of the same domain, where each theory is best for some (scientifically plausible) goal, but where the two theories posit incompatible ontologies. That is, it is possible for us to have goal-dependent pluralism in our scientific ontologies. This ontological pluralism arises simply from our inability to directly know the world's objects, rather than any particular claims about our cognitive limits, values, or social structures. I then present two case studies in which this possibility actually occurs—one based on simulations and theoretical analyses of constructed causal systems, and one from actual scientific investigations into the proper ontology for ocean regions.

Keywords Scientific ontology · Pluralism · Goals · Simulation · Causal prediction · Ocean index

1 Theories, goals, and ontology in science

A widespread maxim among (scientific) naturalists about ontology says (roughly): "the world contains whatever our best sciences say it contains." This maxim is ambiguous, however, about what counts as the "best" sciences. One common reading is that the only relevant science is physics, and in fact, only the part of physics dealing with "fundamental constituents of matter" (whatever those might be). This specific version

D. Danks (✉)
Departments of Philosophy & Psychology, Carnegie Mellon University,
161 Baker Hall, Pittsburgh, PA 15213, USA
e-mail: ddanks@cmu.edu

Published online: 07 January 2015

of the maxim is typically motivated by some combination of the strong empirical success of physics and a reductionist view of the sciences in which physics is the "most basic," and so most fundamental, science. There are many different reasons to worry about this strongly reductionist view, however.[1] I thus take as my starting point the idea that sciences other than physics can tell us something about ontology; that is, the objects of other sciences are not "mere" mereological sums (e.g., my left thumb + my dog's right ear), but objects in their own right.[2] For example, if our best ecological theory holds that there "really are" penguins, then we (*qua* philosophers) should respect that judgment and ought to accord a penguin more ontological status than other mereological sums (though see Chakravartty 2013).

The discussion in the preceding paragraph assumed that we had some way of determining, for a particular domain or phenomenon, which scientific theory is the "best," at least for ontological questions. In actual scientific practice, however, there is no oracle that can tell us which theory is "best." Instead, we must use a variety of different evaluation methods—both quantitative (e.g., statistical tests, Bayesian inferences, mathematical model fit between the world and the theory, or other measures of theory properties) and qualitative (e.g., conceptual consistency with theories in "close" domains, breadth of explanatory power, or range of applicability)—that are based on particular theoretical virtue(s). Determination of which theory is "best" is a highly complex matter; we cannot simply provide a set of theories and some data to a black box that can definitively answer the question for us.

In particular, I claim that whether theory T_1 or T_2 is "best" can sometimes depend in part on the evaluation methods or standards that are used. Two case studies will be provided in Sect. 3, but the general point should be a familiar one to most philosophers of science. For example, a simple-but-false theory *SF* might provide more accurate predictions than a complex-and-true theory *CT* (Geman et al. 1992; Kitcher 2001; Vapnik 1995), and so which is "best" depends on whether evaluation is based on truth (so *CT* is best) or predictive accuracy (so *SF* is best). We are thus naturally led to ask: what is the "right" evaluation standard? That is, given that there are multiple ways to evaluate scientific theories, which one(s) ought we *qua* scientists adopt? I argue in Sect. 2 that there is no single, univocal answer to this question, even if we are principally interested in ontological questions; there is no privileged evaluation method to determine which theory and ontology is "best." Instead, we have multiple goals, and so (potentially) multiple theories of *P* that all can lay claim to being the "best" (even in the unreachable infinite limit of inquiry).

Of course, the possibility of multiple "best" theories for some phenomenon or domain is not in itself ontologically problematic. In global climate modeling, for example, different models will be best depending on whether the goal is short-range or long-range prediction (Solomon et al. 2007), but those different models do not fundamentally differ in their ontologies. The "different" objects in the different models

[1] A (non-exhaustive) sample of concerns can be found in, e.g., Dupré (1993), Cartwright (1999), Kitcher (2001), and Horst (2007).

[2] Of course, the "true" commitments of a scientific theory often cannot be determined solely from the mathematical or linguistic expression of the theory, but instead require further analysis. None of what I say in this paper will turn on this particular difficulty, though.

arise from differences in granularity, rather than any deep incompatibility. That is, the models agree about the underlying ontology; the models for long-run prediction simply aggregate (for computational reasons) some of the objects in the short-run prediction models. More generally, pluralism about theory use is arguably widespread: square pegs and round holes can be modeled as collections of atoms or rigid objects (Putnam 1975), behavior can be generated by groups of neurons or beliefs and desires (Dennett 1987), and so forth. Such pluralism is entirely consistent with monism about ontology when the different theories are at different levels of description or analysis.

Unfortunately, matters are not always so simple: pluralism about "best" theories can yield ontological pluralism (Kellert et al. 2006). In particular, the case studies in Sect. 3 provide examples in which (i) scientific theories T_1 and T_2 have the same domain and level of analysis; (ii) each is "best" according to some scientifically respectable, ontologically relevant evaluation standard; but (iii) they posit truly incompatible ontologies. That is, the (scientific) naturalist about ontology can potentially find herself forced into ontological pluralism. Crucially, this particular ontological pluralism does not arise because of incompatibility between theories about different domains (Cartwright 1999; Giere 1999; Wilson 2006), but rather because of incompatible ontologies in theories within a single domain. This "goal-dependent ontological pluralism" holds that multiple, incompatible ontologies are (potentially) possible, even within a single domain or for a single phenomenon. Moreover, this type of pluralism arises from the mathematics of the scientific theories and relevant empirical domains, rather than from sociological, cognitive, or more pragmatic factors. It is sometimes thought that pluralism arises only because of the "messiness" of interactions within scientific communities, or because of the cognitive limitations of particular scientists. In contrast, this paper argues that ontological pluralism can arise simply from our inability to directly observe the world's "true" ontological structure, combined with our multiple mathematical criteria for when we "understand the world."

It is important to be clear about the nature of this ontological pluralism. The claim is that which parts of the world "count" as objects (rather than being parts-of-objects, or "mere" mereological sums of objects) can change depending on our goals or evaluation criteria. The world does not come to us "carved up" into its constituent objects, and we can "segment" the world (through our scientific theories) in different, incompatible ways depending on our goals and purposes. Moreover, these differences in how to segment the world are meaningful in a way that we (*qua* philosophers) should respect. This ontological pluralism is not the absurd view that objects somehow pop into and out of existence depending on our goals, nor am I advancing an idealist conception of objects. Rather, the position is that there are multiple incompatible ways that we ought to understand our world depending on our goals, and so the pursuit of a unified (scientific) ontology is fundamentally misguided.

Ontological pluralism based on incompatible scientific theories is certainly not novel, but the present paper offers a new argument for this old(er) position.. The most closely related prior position is Dupré's (1993, 2012) promiscuous realism that holds that there are "irreducibly different kinds of things" (Dupré 1993, p. 92), in large part because questions about the natural kind to which an object belongs "can be answered only in relation to some specification of the *goal* underlying the intent to classify the object." (Dupré 1993, p. 5, emphasis added) That is, Dupré argues that we get fundamentally different pictures of the kinds of things in the world depending on our goals, and so (given the multiplicity of our goals) we have ontological pluralism.[3] This multiplicity of goals arises for many reasons, including the fact that scientists are humans with their own personal interests, values, and uses for science, as well as our cognitive inability to understand the world in its full complexity.[4] In contrast, the present paper proposes a more minimal basis for the diversity of goals, coupled with empirical demonstrations that this resulting diversity leads to ontological pluralism. We need not appeal to our human peculiarities or cognitive limits; ontological pluralism is arguably almost-inevitable for any scientific agent that cannot directly observe the "joints" of the world.

Horst's (2007, 2011) Cognitive Pluralism gives a different set of arguments for pluralism, as he holds that we carve up the world in multiple, incompatible ways because of "the ways minds like ours represent features of the world," (Horst 2007, p. 5) including the variety of goals and purposes that we have. Horst argues that different models can be more or less "apt" (to use the term from Horst 2011) depending on our (cognitive) goals, and so ontological pluralism can result. Importantly, Cognitive Pluralism is not based solely on our cognitive limits; its pluralism is not because the world is necessarily too complex for us, but rather because of our modes of engagement with, and understanding of, the world. Nonetheless, this position is strongly cognitivist, and thus idealist/Pragmatist (in the spirit of James, Peirce, and Dewey), while the position I advance in this paper does not require an essential focus on the cognitive (though it is certainly consistent with that additional commitment).

Another common route to ontological pluralism, found in Dupré, Horst, and other pluralists, proceeds through more general arguments against reductionism. More precisely, suppose we think, as the naturalist about (scientific) ontology presumably does, that we are ontologically committed to whatever objects are required to (scientifically) explain the relevant phenomena. If reductionism is false for theory T in some domain, then we are committed to the existence of the objects posited by T. That is, if we cannot reduce the T-objects to objects in some "lower" theory, then we must allow for them to have "real" existence.[5] And so if reductionism is false, then one "only" needs to show that there are equally "good" theories with incompatible ontologies (e.g., for 'species' in Dupré 1993; though see Wilson 1996) to yield ontological pluralism. In contrast, my argument is not about reductionism *per se*, but rather focuses on the necessary multiplicity of evaluation criteria and goals, coupled with a commitment to the ontological legitimacy of the objects of our best sciences.[6] Instead of arguing that we cannot break objects into components without loss (i.e., we cannot go

[3] Kitcher (2001) makes a similar argument, but remains open to the possibility that there could be a single, unifying ontology that works best for all goals and purposes.

[4] Thanks to an anonymous reviewer for emphasizing the many different sources of goals, and so pluralism, for Dupré.

[5] There are numerous complexities here, such as distinguishing clearly between a reduction of a theory and a reduction of the objects of that theory. Dupré (1993) and Horst (2007) each work through these issues in significant detail.

[6] I thus concede that my argument may not have much influence on those who believe in a hard-line eliminativism that says every "higher-level" object is necessarily just a mereological sum. I briefly return to this issue in the Conclusion.

down from the "higher" levels), I will argue in the other direction: we can assemble those components into ontologically respectable (for the scientific naturalist) "higher-level" objects in multiple, incompatible ways. Put more colloquially, there are many different, acceptable, incompatible ways to move up from "lower" levels. Ontological pluralism does not require grand arguments about reductionism *in toto*, or about possibly insuperable difficulties of unifying large scientific theories; it emerges even at the small-scale because of our multiplicity of goals, purposes, and uses—and so multiplicity of evaluation methods—for scientific theories.[7]

One final caveat is in order: I will largely avoid discussion of many recent positions and debates in analytic metaphysics and ontology (e.g., Sider 2011 and the many debates that followed about that book; Wilson 2008), both for lack of space and for more principled reasons. Many of those debates have focused on the nature of "fundamental objects" or "natural properties," where both are understood as quite low-level things (e.g., "electron" or "charge"). In contrast, I am focused on higher-level things; as I noted at the outset, I am interested in sciences other than physics.[8] A different set of analytic metaphysics and ontology debates are potentially relevant, namely those that center on the nature of constitution, composition, and grounding—e.g., if X is composed of a bunch of Y's, then is there *really* an X there or just the collection of Y's? I contend, however, that this connection is largely only apparent. As noted earlier, I am presupposing that the "objects" used in our best (higher-level) sciences are worthy of *some* kind of ontological deference, and then asking what follows from this stance. I am not making the stronger assumption that the things in our best sciences are "really" objects (whatever exactly that is supposed to mean), but only that the scientific naturalist is committed to their having *some* measure of ontological status beyond "mere" mereologial sums. And that minimal commitment is, given the arguments in this paper, sufficient to yield a type of ontological pluralism that is inconsistent with many monist positions in the philosophy of science, though it is perhaps (depending on the degree of ontological deference that should be accorded the sciences) consistent with analytic metaphysical/ontological eliminativism about higher-level objects (cf. footnote 6).

2 A privileged goal?

Discussions of the role of goals or purposes in science often adopt a stance in which goals are described and understood in qualitative, somewhat nebulous, terms. This is perhaps unsurprising: we do not have good mathematical theories of what it means to have a goal, and scientists themselves frequently leave their goals only implicit in

[7] I will also focus more narrowly on scientific contexts, largely because the arguments in the next section require a level of specificity that often does not arise in everyday or commonsense reasoning. Both Dupré (1993) and Horst (2007) argue that (their forms of) ontological pluralism arise even for everyday concepts.

[8] To be clear: I think that many of the arguments in this paper apply equally well to both physics and "higher-level" sciences. It is sometimes argued in present-day analytic metaphysics and ontology, however, that investigations into the "fundamental" or "foundational" constituents of nature can proceed through other means (e.g., Sider 2011). I aim to remain relatively agnostic about those debates, and so prudence leads me to not discuss physics in any significant detail.

their research and writings. In an effort to be at least somewhat more precise, we can follow scientific practice and operationalize the goal in terms of one or more evaluation functions. That is, we judge that we have met a goal just when we are successful relative to one or more evaluation functions. Crucially, we are not *identifying* the goal with the evaluation function(s), but simply observing that we can only assess performance relative to a goal by using such function(s). Goals are valuable precisely because they give us a target towards which we strive. In fact, there is growing evidence that goals play a role in all human learning, including everyday learning outside of scientific contexts; people often efficiently learn only what is required to satisfy their goals—immediate and long-term—rather than trying to learn everything about some domain (e.g., Danks 2014; Wellen and Danks 2014). That being said, if we have no way to evaluate our progress towards or away from some goal, then it is hard to see how that goal can be efficacious in any way. The target provided by a goal loses its value if we are unable to determine our "movement" relative to it. A goal is simply a slogan if we have no means of assessing where we stand relative to it.

Consider now whether there is a single goal, and so a privileged evaluation function for scientific investigations of the world's ontology. If there is really only one (ontological) goal in science, then ontological pluralism presumably cannot arise. More generally, one common view of science is that it has the single, privileged goal of "finding the truth" or "determining what the world is like." All uses of scientific theories for other purposes are then essentially applications of science, not science itself; the goal of the scientific enterprise is "just" to understand and explain our world. Of course, this is a stunningly hard challenge, but nonetheless a single one. If there is a single, objective world and science has the single goal of understanding that world, then science will ultimately deliver a single ontology. There could be short-run "pluralism" due to short-run errors in our scientific theories, but those should (on this picture) eventually be resolved. Different uses of science could subsequently appear to have incompatible "ontologies," but those would presumably arise only because the different scientific applications involve different approximations to the one true (scientific) ontology, rather than any real incompatibility. This picture of scientific goals and ontology is rarely presented in such direct language, but is implicit in much of the way that people talk about issues such as inter-level theoretical reduction (as noted by Dupré 1993; Horst 2007). For example, efforts to reduce cognitive psychology to neuroscience only make sense if there is a single, underlying ontology, presumably that of neuroscience, on which both theories are (or could be) based.

The core issue is whether "understand the world" is actually a single, meaningful, epistemic goal (whose answer would have clear ontological content). The problem is that there does not seem to be an evaluation method or methods for the goal of "understand the world," given that we do not directly observe the world's ontology. One might suggest that theory T_1 is better than theory T_2 just to the extent that T_1 better captures or describes the objects and relations that actually occur in the world. However, actual science is not conducted at some Archimedean point from which we can directly assess whether our scientific theory mirrors the structure of the world (Kitcher 2001). This proposed simple evaluation standard assumes that we can compare our theories against the "ground truth," but that ground truth is exactly what our scientific investigations are trying to establish. We can only directly assess our progress towards this purely epistemic "goal" if we have already reached the goal.

Since we cannot directly establish that we have understood some piece of the world, the ontology-revealing goal will be meaningful only if we can make indirect evaluations relative to it. The strategy of looking for converging evidence in science is, I suggest, exactly a way to make such an indirect evaluation. We do not think that the sciences are telling us something about the structure of the world because we can directly check whether the ontology of a scientific theory is right, but rather because the relevant theory makes accurate predictions, supports the design of experiments and interventions, provides explanations of other phenomena, and so forth. The core intuition here is, of course, the same as the one underlying the No Miracles Argument. The proponent of the single, purely epistemic goal contends that we can use these different evaluation methods and standards to indirectly assess our progress towards understanding the world.[9] Moreover, I would suggest that this type of indirect assessment is entirely standard in the sciences; the proponent of indirect evaluation of the purely epistemic goal correctly captures many of the intuitions behind various scientific practices.

The shift from direct to indirect evaluation of progress towards that goal is a deeply significant one, however. In particular, this move introduces a new assumption that can be empirically tested. Indirect evaluation will only work if the truth (whatever it may be) is always best when it comes to the various evaluation methods (given that we ignore sampling variation, measurement error, and so forth). The whole idea of converging evidence is that the theory—more specifically in our case, the ontology—that is best along one evaluation dimension will also be best along the other dimensions (again, leaving aside issues such as theory T being, say, more accurate than the truth at retrodictions, simply because T tracks sampling variability).[10] We already have significant theoretical reason to worry that the truth is not always best. For example, there are cases in which models that are simpler than the (postulated) truth provably are better predictors than the truth for any finite sample (Vapnik 1995); that is, ontological truth and predictive accuracy provably sometimes diverge given finite data. And as argued above, we cannot determine the ontological truth directly, but rather must infer it from success on indirect evaluation functions.

Of course, perhaps our world is "nice" in the sense that the truth really is best according to the many different (indirect) evaluation functions. This is an empirical question: the claim that the truth has a special status is contingent, and can be empirically tested. We might well have the intuition that "the truth is always best," but there

[9] There are many long-standing philosophical objections to inferring the existence of As from the diverse and repeated successes of a theory that uses As in some essential manner (e.g., Laudan 1981; van Fraassen 1980). If one embraces those objections, however, then one has already given up on the purely epistemic goal, and so will need no convincing that science can have multiple, ontology-relevant goals.

[10] A different proposal would be to argue that the truth will be best according to some weighted combination of the different evaluation dimensions. This would yield a single evaluation function, and so the danger of pluralism would not rear its head. I do not know of any concrete proposals of this form, however, nor is it clear what grounds could be provided to motivate any particular weighting. At the least, it seems that any proposed weighting would likely exhibit significant context- and domain-dependence. Nonetheless, it remains a possibility that could be explored. Thanks to an anonymous referee for emphasizing this point.

is no necessary reason that it should be so. Instead, we must investigate to see whether the true ontology is best for all (scientific) purposes, rather than simply assuming it *a priori*. Importantly, we have reached this point without making any substantive assumptions about the social structure of science, cognitive limits of scientists, or the pragmatic desires and needs of particular people. The argument for a multiplicity of ontology-relevant goals requires only that we be unable to directly observe the world's ontology. The claims in this section are not about the particular kinds of scientists that we happen to be, but rather about more general limits on direct scientific investigation. Moreover, I attempt to show in the next section that, at least in worlds such as ours, the truth is unfortunately not always best.

3 Case studies of incompatible ontologies

This section presents two case studies of incompatible ontologies—two situations in which theories T_1 and T_2 posit different things in the world, and where which theory is "better" depends on which of two (scientifically plausible) goals and evaluation standards is used. The case studies come from different domains and involve different challenges, but have important similarities. In both instances, we start with "low-level" data over a set of variables $\mathbf{R}$, and then face a scientific challenge to define, based on $\mathbf{R}$, a set of "high-level" variables $\mathbf{C}$, where we think that $\mathbf{C}$ is the "right" way to characterize the system. That is, part of the scientific challenge in each case study is precisely to determine the "right" ontology for the system. This challenge arises in many different domains; for example, neuroimaging measurements (e.g., in an fMRI experiment) are often more spatially fine-grained than (what we think are) the "real" brain regions, and so we must construct variables for the so-called regions of interest from those fine-grained measurements. The interesting result in each case study below is that different high-level ontologies are each "right" for some scientific goal. Of course, there is agreement about the underlying $\mathbf{R}$ data, so the door is open to a certain type of ontological eliminativism that denies that there is *anything* other than $\mathbf{R}$. As noted in Sect. 1, though, I presume that things above the level of fundamental physics sometimes deserve some measure of ontological deference. The problem is that incompatible objects—that is, sets $\mathbf{A}$ and $\mathbf{B}$ such that we cannot translate from descriptions in terms of $\mathbf{A}$ to descriptions in terms of $\mathbf{B}$, or vice versa—seem to each deserve that same deference.

3.1 Different ontologies for different types of predictions

In many scientific settings, we are interested in predicting some key target variable T. For example, we might want to determine which students in an online course are likely to pass the final exam, or which individuals in a population are likely infected with a disease, or which behavioral changes might lead to weight loss. Although we presumably measure many different variables, there is one variable that is of particular importance in each case. At the same time, there are at least two different types of predictions that we might want to make: prediction from observation or prediction from intervention. In the former case, we want to predict something about T's value from observations of other features of an individual or case. In the latter case, we want to predict T given that there is an intervention (from outside of the system) to set or change some of the other variables in the system.[11] It is easy to see that these two types of prediction can be quite different. For example, my prediction about the state of a light switch will change depending on whether I observe the lights to be off, or instead force them to be off by smashing them. In the former case, I infer that the light switch is probably in the "off" position; in the latter case, no such inference is warranted. Alternately, if T is the job title of an academic (e.g., Assistant vs. Associate vs. Full Professor), then the prediction for an individual observed to have gray hair is presumably quite different than for an individual whose hair is dyed gray.

The general point is that predictions depend in part on the causal structure of the system, and interventions can change the causal structure in important ways. Suppose that the true underlying causal structure is $T \to E$; that is, T is a direct cause of E. In general, observations of E will be informative about the state of T, as when observations of symptoms are informative about whether someone has a disease. Interventions to force E to have a particular state, however, will break the $T \to E$ connection, and so make E uninformative about T (since T now plays no role in determining E's value). Interventions do not always change the predictions, though. Suppose instead that the underlying causal structure is $C \to T$; that is, C is a direct cause of T. In this situation, we will make the same predictions regardless of whether we observe C's state or intervene to produce that state in C.[12] For example, we can (in normal conditions) make the same predictions about the state of the light bulb regardless of whether we observe or intervene the switch to be "on." In general, observation is symmetric—observing either causes or effects of T will be predictively useful—while intervention is asymmetric—only intervening on causes of T (but not its effects) will be predictively relevant (as noted by many authors, including Hausman 1998; Pearl 2000; Spirtes et al. 1993; Woodward 2003). Thus, prediction from interventions requires a level of knowledge and specificity about the underlying causal structure that is not required for prediction from observations.

Both prediction from observations and prediction from interventions are significant scientific goals, and one might naturally wonder whether they can yield different ontologies. More specifically, suppose we have a target T and a set of other, "low-level" variables $\mathbf{R}$. We can construct a set $\mathbf{C}$ of "high-level" variables that are deterministic functions of $\mathbf{R}$; for example, perhaps $C_3 = R_5 + R_{17} - R_{28}$. In general, we can allow for the constructed $\mathbf{C}$-variables to be quite sophisticated functions of the $\mathbf{R}$-variables. Define $\mathbf{C_{Obs}}$ to be the optimal "high-level" variables for predicting some relevant aspect of T given observations; that is, $\mathbf{C_{Obs}}$ is the set of variables such that predictions of T given observations of $\mathbf{C_{Obs}}$ are better than predictions of T given

[11] And of course, for each type of prediction, there are many different aspects of T that we might want to predict, such as its precise value, expected deviation from some norm, variance over time, and so forth. I briefly return to this issue below.

[12] Assuming that various technical conditions hold, such as (i) we have an ideal intervention; (ii) there are no common causes of C and T; etc.

observations of any other set of constructed variables.[13] Similarly, we can define $\mathbf{C_{Int}}$ to be the optimal "high-level" variables for predicting some relevant aspect of T given interventions on $\mathbf{C_{Int}}$. We can now restate the problem as: is it always the case that $\mathbf{C_{Obs}} = \mathbf{C_{Int}}$?[14] If we think that "the truth is always best," then we should expect the two optimal sets to always be the same (at least, in the limit of infinite $\mathbf{R}$-data). This intuition turns out to be incorrect, at least when we place ourselves in a realistic epistemological setting.

Fancsali (2013) examined this question through extensive simulation studies, and a subsequent theoretical analysis. Although there are obvious ways in which simulations are artificial, they can also be valuable tools. In particular, we can fully control the (simulated) causal structure, and so can both know the "truth" and also examine why we get the results that we do. In particular, we can ask, for some given causal structure over $\mathbf{R}$, whether $\mathbf{C_{Obs}} = \mathbf{C_{Int}}$, both given finite data and in the limit of infinite $\mathbf{R}$-data. Fancsali focused on causal structures in which the target variable T really is caused by higher-level constructs of the measured raw variables, and so there is a non-trivial higher-level ontology to be discovered. That is, there is something like a "truth" that we might hope to discover using diverse indirect measures; we can directly test whether "the truth is always best," at least in this simulated world. Crucially, Fancsali made one further, realistic assumption: for given $\mathbf{C}$, we do not know the causal structure *a priori*, but rather must learn it from our data. Causal learning from data is a non-trivial (but not impossible) learning task in this situation (Chickering 2002; Spirtes et al. 1993). For certain $\mathbf{C}$-variables, however, there could be uncertainty about the underlying causal structure,[15] and so the evaluation functions—the measures of predictive power given observations or interventions—had to accommodate this uncertainty. Fancsali used the expected predictive power, but his qualitative findings were not sensitive to that choice.

Across a range of simulation settings, Fancsali (2013) found that $\mathbf{C_{Obs}}$ and $\mathbf{C_{Int}}$ frequently differed. That is, the high-level constructed variables that were best for predicting T given observations were different from the high-level constructed variables that were best for predicting T given interventions. There are many technical details that I omit here, though see Fancsali (2013) for details. For example, there are multiple targets of prediction given interventions (roughly, precision vs. magnitude), and so multiple $\mathbf{C}$-sets for a given dataset. Also, one must be careful to specify exactly how a $\mathbf{C}$-variable intervention is instantiated in interventions on the $\mathbf{R}$-variables, as there will typically be multiple $\mathbf{R}$-level realizations that have the same $\mathbf{C}$-level impact. Nonetheless, across many different ways of specifying these details, the truth was *not*

13 One might object that the relevant standard should be given observations of $\mathbf{R}$, not some other $\mathbf{C}$. However, we can allow for $\mathbf{C}$ to be the trivial functions of $\mathbf{R}$ (i.e., $\mathbf{C} = \mathbf{R}$), and so include this possibility.

14 As noted earlier, we can actually have multiple $\mathbf{C_{Obs}}$ and $\mathbf{C_{Int}}$ sets depending on exactly what we want to predict about T. In the analysis described below, for example, Fancsali (2013) considered multiple criteria for $\mathbf{C_{Int}}$, including "predicting post-intervention magnitude of T" and "predicting post-intervention change in T." The emergence of incompatible ontologies did not depend on the particular aspect of T that one wanted to predict.

15 For example, if we know only that C and T are correlated, then we do not know whether $C \rightarrow T$ or $C \leftarrow T$ (or perhaps a common cause of the two). We know something about the causal structure—namely, there is some causal connection—but still have some uncertainty.

always best: there was no single set of variables—in particular, not the variables actually used to generate data—that was optimal for both prediction given observations and prediction given interventions.

Moreover, the divergence between $\mathbf{C_{Obs}}$ and $\mathbf{C_{Int}}$ was not due to sampling variation in finite data, but arose for principled theoretical reasons that hold even in the limit of infinite data. The underlying issue is that there can be a (hidden) trade-off between (a) predictive power; and (b) certainty about causal structure. In general, our ability to learn the causal structure around T depends on not only the associations between T and other variables, but also on the associations between those other variables.[16] For example, suppose I know that A and B are each associated with T both unconditionally and conditional on the other (and for simplicity, assume that there are no unobserved common causes). This information implies that A and B are each directly causally connected to T, but tells us nothing about the orientation of those direct causal connections. To learn more about the orientations, we need to know how A and B are related to each other. For example, if they are unconditionally associated (and independent conditional on T), then the causal structures consistent with these observations are: $A \rightarrow T \rightarrow B$; $A \leftarrow T \rightarrow B$; and $A \leftarrow T \leftarrow B$. In contrast, if A and B are unconditionally independent, then the only causal structure consistent with these observations is: $A \rightarrow T \leftarrow B$. Importantly, our predictions given interventions will typically be worse in the former case than in the latter one, precisely because we are uncertain about the underlying causal structure. There is, for example, no determinate prediction given an intervention on A, since A is a cause of T in one possible structure but not the other two.

In general, predictions given observation depend only on the association structure between the proposed $\mathbf{C}$-variables and T, while predictions given interventions also depend on the associations among the $\mathbf{C}$-variables (since those influence our uncertainty about causal structure). This additional constraint on $\mathbf{C_{Int}}$ implies that we can get divergence from $\mathbf{C_{Obs}}$ even in the limit of infinite data, as the constructed variables that are most highly associated with T in the limit of infinite data need not have a suitable association structure with one another to reduce causal structure uncertainty. The cases that Fancsali (2013) explored have exactly this structure: the high-level variables that best predicted T, even in the limit of infinite/perfect data, often also yielded substantial causal structure uncertainty. In contrast, there was sometimes a different set of high-level variables that was slightly worse at predicting T (i.e., the variables were slightly less associated with T) but provided much more certainty about the causal structure (i.e., the between-variable association structure was better for causal learning). For prediction given observation, there is no incentive to engage in this trade-off: uncertainty about causal structure just does not matter. But for prediction given intervention, there can be a significant advantage to having increased certainty

[16] At least, if we make some general assumptions about how causal relations manifest in observed data. I do not engage here with the debate about the legitimacy of those assumptions (see, e.g., Cartwright 2002, 2007; Glymour 1999), as this example only requires that we can sometimes learn causal structure without experiments. Even opponents of the general project of causal structure learning from observations grant that the assumptions are sometimes acceptable.

about causal structure, and this advantage is sometimes sufficiently great that it can offset the loss of power in prediction from observations.

In terms of the broader issues discussed in this paper, the key is that the two sets of high-level, constructed variables—those that are optimal for prediction given observation vs. prediction given intervention—provide incompatible (high-level) ontologies. In this context, two ontologies $\mathbf{C}_1$ and $\mathbf{C}_2$ are incompatible just when their respective variables are based on different **R**-variables. For example, if $C_1 = f(R_{12}, R_{17})$ and $C_2 = g(R_3, R_{68})$, then the ontologies are incompatible; in contrast, if $C_1 = h(R_8, R_{27})$ and $C_2 = k(R_8, R_{27})$, then although the functions are different, the high-level variables are arguably ontologically compatible since they are based on the same **R**-variables. In Fancsali's (2013) simulations and subsequent theoretical analyses, however, $\mathbf{C_{Obs}}$ and $\mathbf{C_{Int}}$ typically contained high-level variables that were based on quite different low-level **R**-variables. That is, different scientific goals can lead to different, incompatible (scientific) ontologies. Moreover, this ontological pluralism arises because of the mathematics of the situation, combined with the minimal epistemological assumption that we cannot directly observe causal structure. We do not need any strong claims about the nature of human scientists or scientific communities; pluralism arises just from the statistical structure of the relevant **R**-variables.

3.2 Different ocean indices for different properties

One might justifiably wonder whether the particular causal structures explored in Fancsali (2013) are realistic, or whether those kinds of situations ever arise in actual scientific practice. This section presents an actual scientific case study in which different, incompatible ontologies resulted from scientists pursuing different goals. Importantly, the different goals are not just different ways we might want to use a scientific theory, but rather arise from our (apparent) use of multiple criteria in everyday object individuation and understanding. Those criteria typically yield a single, univocal answer about how to divide the world into objects, but in this case study, the multiple corresponding scientific goals yield incompatible ontologies. The individuation criteria do not converge on a single way of understanding the world.

Water temperature and atmospheric surface pressure vary across the oceans, but not in a purely random manner. In particular, large areas of the ocean can behave relatively uniformly. For example, the El Niño phenomenon involves a large area of the Pacific Ocean off the South American coast becoming particularly warm in a relatively coherent manner. These large ocean areas can have long-range impacts, called teleconnections, on both terrestrial weather patterns and other ocean regions (Glantz et al. 1991), such as El Niño typically causing increased snowfall in the U.S. Rocky Mountains (Ropelewski and Halpert 1996). A natural scientific question is: what is the proper ontology for these large ocean regions, also called ocean indices? That is, how ought we "carve up" the ocean surfaces into regions that should be regarded as distinct "objects"? The ocean obviously does not come to the scientists neatly divided; rather, they must learn the high-level ocean index ontology from observational data.[17] In particular, scientists have measurements of properties of particular ocean locations (e.g., water temperature, atmospheric surface pressure, etc.) from a vast array of measurement buoys. Those measurements must then be aggregated into ocean indices for large ocean areas; for example, scientists might conclude that buoys #782 through #1,204 constitute measurements of a single, high-level ocean index. Of course, it is possible that there is no stable ocean index ontology, but that should be empirically determined, not assumed *a priori*.

I introduced ocean indices by pointing towards two different aspects of them: they are regions that (a) exhibit substantial internal coherence; and (b) can serve as relata in stable, long-range climate teleconnections. These different aspects suggest different evaluation methods for scientists attempting to determine the high-level ocean index ontology. On the one hand, scientists might look for ocean regions such that buoy measurements within that region are strongly correlated with one another. This method is analogous to segmenting the everyday world into objects based on high correlations in location; part of why the handle of the coffee cup on my desk is part of the "mug object" rather than the "computer object" is because the handle's location is strongly correlated with the locations of other parts of the "mug object," but not with the locations of parts of the "computer object." In general, high-level objects typically exhibit strong internal correlations along some dimensions, and this evaluation method attempts to discover ocean index objects using that feature. On the other hand, scientists could look for ocean regions that stand in stable (over time) causal relations with one another, and with various terrestrial weather phenomena. This method is analogous to picking out objects in the everyday world based on which parts of the world interact causally (and stably) with which other parts. For example, moving the handle of my coffee cup reliably causes a change in the mug part, but not a reliable change in my computer.

In the everyday world, these two methods for object individuation are highly correlated, though conceptually distinct. Strong internal correlation and stable causal relations tend to pick out the same objects, but there is no (obvious) necessary reason that this must be the case. In the case of ocean indices, the criteria come apart: these two different evaluation methods—alternately, these two different scientific goals—lead to incompatible high-level ontologies. Segmentation based on internal coherence and correlation leads to a stable set of ocean indices $\mathbf{O_{ICC}}$ that significantly outperforms the ocean indices proposed by oceanographers and climatologists, $\mathbf{O_{SOI}}$, on a wide range of prediction tasks (Steinbach et al. 2003). In terms of stable causal relations, however, $\mathbf{O_{SOI}}$ is clearly superior to $\mathbf{O_{ICC}}$ (Chu and Glymour 2008; Glymour 2007), as causal relations involving the latter are sensitive to statistical artifacts that ought (on any standard understanding of causality) not change the causal relations. Crucially, neither set of ocean indices is a more fine-grained version of the other; $\mathbf{O_{ICC}}$ and $\mathbf{O_{SOI}}$ carve the world up in truly different ways, even though each is derived from a scientifically important goal and criterion. We thus have a real-world example in which we get different high-level (scientific) ontologies depending on whether our goal is to find ocean regions that are internally coherent or (externally) causally stable.

[17] Interventions directly on ocean indices are not technically feasible at the current time.

Incompatible ontologies are not merely theoretically possible, but actually arise in scientific practice.

4 Conclusion

The conclusions of this paper emerge straightforwardly from a set of individually intuitive notions. First, one way to determine the world's ontology is to look to our best sciences, where we accord some degree of ontological status[18] to things even in (higher-level) sciences. Second, the notion of "best" in the preceding sentence is relative to the particular goal or purpose that we have. Third, there is no single, purely epistemic goal (that can be directly assessed) for the sciences. Therefore, it is possible (though not necessary) that we can have multiple "best" sciences with incompatible ontologies for some domain. In Sect. 3, I then presented two case studies to argue that this possibility is actual: sometimes, we have multiple, incompatible ontologies for the very same domain, where the proper ontology depends on one's goal or purpose. This final conclusion, of course, looks much more radical than the individual claims, and suggests that I am advocating some type of strong idealism about the world's ontology. One could certainly turn in that direction (e.g., Horst 2007), but the idea behind this paper is more modest: it is simply the observation that the world does not come to us already carved up into distinct pieces or objects, and the only ways we have to assess possible "carvings up" depend in part on what we want to do with that "carving up" (see also Dupré 1993). Moreover, we arrive at this pluralistic conclusion just by assuming that we do not directly observe the world's ontology; we do not need to invoke general arguments against reductionism or monism, or make reference to specific details about human cognitive limits and values, or the social structure of scientific communities.

Of course, this argument has left open a number of key ontological questions. Most notably, I have aimed to remain agnostic about the exact ontological status of the things in each "best" theory; I assumed only that they deserve some measure of ontological status. One response to my arguments for pluralism would thus be to deny any ontological status at all to the things of any scientific theory other than fundamental physics (or whatever is the "fundamental" theory). This hard-line eliminativism gives up, however, on the possibility that any sciences other than physics tell us something about the structure of the world beyond "just" which clusters of quarks (or strings, or whatever) have properties that we happen to find appealing. Such a position arguably flies in the face of much of scientific practice. And if we do not want to take such an eliminativist stance, then we are arguably led towards ontological pluralism: nature can be carved up (in some ontologically meaningful sense) in many different ways depending on one's ontology-relevant goals.

[18] Again, I aim to bracket off the question of exactly what this ontological status is, and in particular, whether these things must be understood to be *real* objects (in some sense). All I require here is that these things deserve *some* type or degree of status.

Acknowledgments Thanks to participants at the 2013 Ontology & Methodology conference at Virginia Tech, as well as three anonymous reviewers for this journal, for their valuable comments, feedback, and criticisms. This work was partially supported by a James S. McDonnell Foundation Scholar Award.

References

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

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, *53*, 411–453.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Chakravartty, A. (2013). On the prospects of naturalized metaphysics. In D. Ross, J. Ladyman, & H. Kincaid (Eds.), *Scientific metaphysics*. Oxford: Oxford University Press.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, *3*, 507–554.

Chu, T., & Glymour, C. (2008). Search for additive time series causal models. *Journal of Machine Learning Research*, *9*, 967–991.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge: The MIT Press.

Dennett, D. C. (1987). *The intentional stance*. Cambridge: The MIT Press.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge: Harvard University Press.

Dupré, J. (2012). *Processes of life: Essays in the philosophy of biology*. Oxford: Oxford University Press.

Fancsali, S. E. (2013). *Constructing variables that support causal inference*. (Doctoral dissertation). Pittsburgh: Carnegie Mellon University, Department of Philosophy.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*, 1–58.

Giere, R. N. (1999). *Science without laws*. Chicago: University of Chicago Press.

Glantz, M. H., Katz, R. W., & Nicholls, N. (Eds.). (1991). *Teleconnections linking worldwide climate anomalies*. Cambridge: Cambridge University Press.

Glymour, C. (1999). Rabbit hunting. *Synthese*, *121*, 55–78.

Glymour, C. (2007). When is a brain like the planet? *Philosophy of Science*, *74*, 330–347.

Hausman, D. M. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.

Horst, S. (2007). *Beyond reduction: Philosophy of mind and post-reductionist philosophy of science*. New York: Oxford University Press.

Horst, S. (2011). *Laws, mind, and free will*. Cambridge: The MIT Press.

Kellert, S. H., Longino, H. E., & Waters, C. K. (Eds.). (2006). *Scientific pluralism*. Minneapolis: University of Minnesota Press.

Kitcher, P. (2001). *Science, truth, and democracy*. New York: Oxford University Press.

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, *48*(1), 19–49.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Putnam, H. (1975). Philosophy and our mental life. *Mind, language, and reality: Philosophical papers* (pp. 291–303). Cambridge: Cambridge University Press.

Ropelewski, C. F., & Halpert, M. S. (1996). Quantifying southern oscillation-precipitation relationships. *Journal of Climate*, *9*, 1043–1059.

Sider, T. (2011). *Writing the book of the world*. Oxford: Oxford University Press.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., et al. (Eds.). (2007). *Climate change 2007: The physical science basis: Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (1st ed.). Berlin: Springer.

Steinbach, M., Tan, P.-N., Kumar, V., Klooster, S., & Potter, C. (2003). Discovery of climate indices using clustering. In *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 446–455). New York: ACM.

van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.
Wellen, S., & Danks, D. (2014). Learning with a purpose: The influence of goals. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 1766–1771). Austin, TX: Cognitive Science Society.
Wilson, M. (2006). *Wandering significance: An essay on conceptual behavior*. Oxford: Oxford University Press.
Wilson, R. A. (1996). Promiscuous realism. *British Journal for the Philosophy of Science*, *47*, 303–316.
Wilson, R. A. (2008). Material constitution and the Many-Many Problem. *Canadian Journal of Philosophy*, *38*(2), 201–218.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.