

CARTWRIGHT ON EXPLANATION AND IDEALIZATION

ABSTRACT. Nancy Cartwright (1983, 1999) argues that (1) the fundamental laws of physics are true when and only when appropriate *ceteris paribus* modifiers are attached and that (2) *ceteris paribus* modifiers describe conditions that are almost never satisfied. She concludes that when the fundamental laws of physics are true, they don't apply in the real world, but only in highly idealized counterfactual situations. In this paper, we argue that (1) and (2) together with an assumption about contraposition entail the opposite conclusion – that the fundamental laws of physics *do* apply in the real world. Cartwright extracts from her thesis about the inapplicability of fundamental laws the conclusion that they cannot figure in covering-law explanations. We construct a different argument for a related conclusion – that forward-directed idealized dynamical laws cannot provide covering-law explanations that are causal. This argument is neutral on whether the assumption about contraposition is true. We then discuss Cartwright's simulacrum account of explanation, which seeks to describe how idealized laws can be explanatory.

“One source of misunderstanding is the view . . . that a hypothesis of the simple form ‘every P is Q ’ . . . asserts something about a certain limited class of objects only, namely the class of all P 's. This idea involves a confusion of logical and practical considerations: Our interest in the hypothesis may be focused upon its applicability to that particular class of objects, but the hypothesis nevertheless asserts something about, and indeed imposes restrictions upon, *all* objects.” (Hempel 1965, p. 18)

In *How the Laws of Physics Lie*, Nancy Cartwright argues that the fundamental laws of physics don't provide true descriptions of how objects behave in the real world. They either make false claims about real objects or true claims that apply only in highly idealized counterfactual situations. For example, when Newton's law of gravitation ($f = Gm_1m_2/r^2$) is interpreted literally, it is false – the net force acting on a pair of objects almost never has the value specified. In defense of this thesis, Cartwright considers two ways that one might propose to reinterpret this law so that it comes out true:

- (1) Interpret the law as including a *ceteris paribus* modifier (i.e., if there are no forces other than gravity at work, then $f = Gm_1m_2/r^2$).
- (2) Interpret the law as describing a component force (i.e., the force due to gravity $f_g = Gm_1m_2/r^2$).



Cartwright rejects interpretation (2) because she denies the reality of component forces; this has elicited criticisms from several philosophers.¹ The correctness of Cartwright's position on this issue is not the focus of our paper. Rather, we want to assess Cartwright's argument that if we interpret the fundamental laws of physics as in (1), then their antecedents will describe conditions that are almost never satisfied; hence, they won't describe how real objects actually behave. With respect to this argument, Cartwright writes:

Ceteris paribus generalizations, read literally without the '*ceteris paribus*' modifier, are false. They are not only false, but held by us to be false; and there is no ground in the covering-law picture for false laws to explain anything. On the other hand, with the modifier the *ceteris paribus* generalizations may be true, but they cover only those few cases where the conditions are right. For most cases, either we have a law that purports to cover, but cannot explain because it is acknowledged to be false, or we have a law that does not cover. Either way, it is bad for the covering-law picture. (Cartwright 1983, 45–46)

Cartwright's argument can be stated as follows:

- (3) The fundamental laws of physics are true only when appropriate *ceteris paribus* modifiers are attached.²
- (4) *Ceteris paribus* modifiers describe conditions that hold only under ideal situations.
- (5) When the fundamental laws of physics are true, they apply only to objects in ideal (counterfactual) situations.
- (6) Therefore, the fundamental laws of physics don't apply to objects in the real world.³

We will argue that even granting the truth of premises (3) and (4), (5) does not follow (hence, neither does (6)).

Cartwright says that the statement "two bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of their masses" is false unless we attach the *ceteris paribus* modifier "there are no forces other than gravitational forces at work". She writes:

Speaking more carefully, the law of universal gravitation is something like this: If there are no forces other than gravitational forces at work, *then* two bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of their masses. I will allow that this law is a true law, or at least one that is held true within a given theory. But it is not a very useful law Once the *ceteris paribus* modifier has been attached, the law of gravity is irrelevant to the more complex and interesting cases. (Cartwright 1983, 58)

Thus, according to Cartwright, a true law will have the form ' $C \rightarrow L$ '. Cartwright argues that L is true only if the qualifier concerning C is attached to it, but C is almost never satisfied in the real world. Hence, the law is a true conditional whose antecedent and consequent are both false. For this reason, Cartwright concludes that ' $C \rightarrow L$ ' fails to apply to real objects. We now will argue that Cartwright's claims (3) and (4) and a plausible principle concerning contraposition entail that ' $C \rightarrow L$ ' *does* apply to real objects.⁴

Let's consider the laws that Cartwright discusses to see if their contrapositives apply to real objects. The first example is the law of gravitation:

If there are no forces other than gravity at work, then $f = Gm_1m_2/r^2$. ($C \rightarrow L$)

This is equivalent to:

If $f \neq Gm_1m_2/r^2$, then there are forces other than gravity at work. ($\sim L \rightarrow \sim C$)

Cartwright claims that ' $C \rightarrow L$ ' does not apply to objects in the real world because C and L are each false of real objects. However, this means that $\sim C$ and $\sim L$ are each *true* of real objects, so presumably ' $\sim L \rightarrow \sim C$ ' applies to real objects. This leads to the unsatisfactory result that a conditional and its contrapositive, though logically equivalent, nonetheless apply to different things.

The same pattern may be found in another example that Cartwright considers – Snell's law. After claiming that Snell's law is false as it is stated in textbooks, she represents Snell's law as follows:

Refined Snell's Law: For any two media which are optically isotropic, at an interface between dielectrics there is a refracted ray in the second medium, lying in the plane of incidence, making an angle θ_t with the normal, such that: $\sin \theta / \sin \theta_t = n_2/n_1$. (Cartwright 1983, 47)

Cartwright thinks that the condition about optical isotropy is almost never satisfied, so the law does not apply to real objects. However, this means that if a medium is such that $\sin \theta / \sin \theta_t \neq n_2/n_1$, then it is *not* isotropic. Thus, the contrapositive of Cartwright's Refined Snell's Law *does* apply to real objects. If a conditional and its contrapositive apply to precisely the same things, then at least one of these judgments about the applicability of a conditional and its contrapositive must be wrong.

Consider a third example from a different science – the Hardy–Weinberg law in population genetics. It states that:

If no evolutionary forces are at work and the gamete frequency of gene A is p and the gamete frequency of gene a is q (where $p + q = 1$), then the frequencies of the genotypes AA , Aa , and aa are p^2 , $2pq$, and q^2 , respectively. (Sober 1984)

The antecedent of this law is never satisfied, so if we apply Cartwright's argument, we should conclude that this law does not apply to real populations. However, if we look at the contrapositive of this law, we see that it *does* apply to real populations. If we observe that the genotype frequencies *aren't* at their Hardy–Weinberg values, then there *are* evolutionary forces at work.

Here is the general pattern: take any true law in conditional form whose antecedent and consequent are false (according to Cartwright, all fundamental laws in physics have this feature). In such a case, even if the conditional itself *seems* to be vacuous, its contrapositive won't be, since the antecedent and consequent of the contrapositive will correctly describe real objects. If a conditional and its contrapositive apply to the same things, then either both apply to real objects or neither does.

Thus far we've seen that three claims are in conflict: That ' $C \rightarrow L$ ' fails to apply to real objects if C involves an idealization, that ' $\sim L \rightarrow \sim C$ ' applies to real objects, and that a conditional and its contrapositive must apply to exactly the same things. Which of these claims should be abandoned? To begin with, we think it is implausible to deny that a conditional and its contrapositive apply to exactly the same things. Since a conditional and its contrapositive are logically equivalent, they are different verbal formulations of *the same proposition*. Laws, it should be remembered, are supposed to be extra-linguistic entities; Newton's law of gravitation is no more a part of English than it is of any other natural language. If laws are propositions of a certain type, then Cartwright's position on contraposition must be mistaken.

We also find it implausible to deny that ' $\sim L \rightarrow \sim C$ ' applies to the systems of which $\sim L$ and $\sim C$ are true. After all, scientists *use* such contrapositives to reason about real world systems. For example, if *this population* deviates from Hardy–Weinberg proportions, then *it* must be undergoing some evolutionary process. If this is not an example of "applying the contrapositive to a real object", we don't understand what "applying" means.

What follows, then, is that we should reject Cartwright's thesis that laws of the form ' $C \rightarrow L$ ' fail to apply to real systems just because C involves an idealization. We suspect that Cartwright drew this conclusion by focusing exclusively on the argument form 'If C , then L . C . Therefore L .' Since C is false in the real world, this argument form cannot be applied to real objects. The point about *the argument form* is correct, but nothing

follows about *the conditional* itself. For the same conditional also plays a role in a different argument scheme, namely ‘If C , then L . $\sim L$. Therefore $\sim C$ ’, and *this* form of argument *does* apply to real objects.

We recognize that Cartwright may want to contest the claim that a conditional and its contrapositive apply to exactly the same things. Given this, it is gratifying that something like Cartwright’s conclusion can be defended without taking a stand on this question. Cartwright’s main point in advancing her claim about the inapplicability of fundamental laws is to develop a point about explanation: *true fundamental laws do not figure in covering law explanations*. The argument we have in mind concerns forward-directed deterministic dynamical laws – laws that have the form “if C holds at time t , then E holds at $t + \Delta t$ ”. Suppose the C in this law describes idealized circumstances. This means that the forward-directed argument *form* does not apply to real systems, but the backwards- directed argument *form* *does*, as we have explained. If explanation must be causal and if causes must precede their effects, then the backwards-directed argument, though applicable to real objects, cannot provide an explanation of its conclusion; one can’t explain what happens at some earlier time by describing the later state of the system. Recall that Cartwright’s goal was to show that fundamental laws don’t provide covering law explanations. The conclusion of the argument we have presented is that forwards-directed deterministic dynamical laws that describe idealized circumstances in their antecedents cannot provide covering-law causal explanations, regardless of whether these laws are classified as fundamental or derived.⁵

What is the situation with respect to dynamical laws that are probabilistic? To begin with, we note that whereas a conditional and its contrapositive are logically equivalent, “ $\text{Pr}(X | Y) = p$ ” and “ $\text{Pr}(\text{not } Y | \text{not } X) = p$ ” are not. Furthermore, it turns out that if “ $\text{Pr}(E \text{ holds at } t + \Delta t | C \text{ hold at } t) = p$ ” is a law, then “ $\text{Pr}(\sim C \text{ holds at } t | \sim E \text{ holds at } t + \Delta t) = q$ ” rarely is. The reason is that laws must be time-translationally invariant; see Sober (1993b) for discussion. The conclusion we draw is that if “ $\text{Pr}(E \text{ holds at } t + \Delta t | C \text{ hold at } t) = p$ ” is a law in which C involves an idealization, then the following argument form will not constitute a covering-law explanation:

$$\begin{array}{l} \text{Pr}(E \text{ holds at } t + \Delta t | C \text{ hold at } t) = p \\ C \text{ holds } t \\ p \text{ [=====} \\ E \text{ holds at } t \Delta t \end{array}$$

The reason is that the second premise is false. Notice that this point applies, regardless of whether we demand that p be high, as Hempel’s (1965)

inductive-statistical model requires, or allow p to take any value, which is what Salmon's (1984) account permits.

The conclusion of our argument, then, is that forward-directed dynamical laws fail to provide covering-law causal explanations, if the laws in question are deterministic and contain idealizations in their antecedents, and if they are probabilistic and contain idealizations in their conditioning propositions. This argument differs from Cartwright's in three ways. First, ours does not gainsay the assumption that a conditional and its contrapositive apply to exactly the same things. Second, it does not require a distinction between fundamental and non-fundamental laws. And third, our argument is restricted to *causal* explanations. Despite these differences, we believe that our argument captures much of what Cartwright is after.

Cartwright's goal was to show that laws that contain idealizations cannot be used in covering-law explanations. She thinks that this has implications for many theories of explanation, not just Hempel's, and so she uses the expression "covering-law model of explanation" in a very wide sense, and we have followed her in this. As noted above, the phrase also applies to Salmon's (1984) model. But what, then, does talk of the covering-law model actually cover? The argument we have constructed pertains to any theory of explanation that requires the following: (i) the *explanans* must describe the cause(s) of the *explanandum*; (ii) the *explanans* must cite a law; (iii) all of the *explanans* propositions must be true; (iv) the *explanans* explains the *explanandum* by entailing it or by conferring a probability on it. Forward-directed dynamical laws that contain idealizations in their antecedents (or in their conditioning propositions, if they are probabilistic) cannot figure in explanations, if explanations must have these features. Cartwright (1983) proposes a "simulacrum account of explanation" as an alternative to the covering-law approach; the main point of this account is to make room for the fact that idealized laws can be explanatory. However, she provides very few details on how this new model of explanation is to be understood. We take up this problem in what follows. Our proposal will be that some of the explanations that idealized laws help provide satisfy conditions (i)–(iii), but not (iv).

In evolutionary biology, optimality models describe the value of a trait that maximizes fitness, given a set of constraints. For example, the optimal length of a bear's fur might be modeled as a function of the ambient temperature, the bear's body size, the energetic cost of growing fur, and so on. These models are often interpreted dynamically – if organisms are fitter the closer they are to the specified optimum, and if natural selection is the only force acting on the population, then the optimal trait value will

evolve. Understood in this way, optimality models contain idealizations; they describe the evolutionary trajectories of populations that are infinitely large in which reproduction is asexual with offspring always resembling their parents, etc. (Maynard Smith 1978; Sober 1993a).

We want to argue that optimality models are explanatory despite the fact that they contain idealizations. As just noted, these models are interpreted as entailing conditionals of the following form:

- (7) If organisms are fitter the closer they are to the optimal value α and if no forces other than selection are at work in the population, then the population will evolve to a state in which all organisms exhibit the trait value α .⁶

Suppose the optimality model correctly describes how selection acts on the trait of interest:

- (8) Organisms are fitter the closer they are to the optimal value α .

Given this information, suppose we observe that

- (9) The n organisms in the population have trait values $\beta_1, \beta_2, \dots, \beta_n$ (where each β_i differs only negligibly from α).

Our question is – do (7) and (8), if true, together explain (9)? We think that the answer is *yes*, even if one can provide no details about the nonselective forces that happen to be acting on the population, and no idea how that more complex situation ought to be modeled.⁷ Proposition (8) provides a partial description of the initial conditions and proposition (7) provides an idealized model whose antecedent applies to no real world system. Granted, these propositions do not constitute a *complete* explanation of (9) in which *all* causally relevant factors are described, but we think they are explanatory nonetheless.

The pattern here is hardly unique to evolutionary biology. Consider the law of gravitation, understood, as Cartwright says it should be, as describing the net force that would be present if gravitation were the only force at work. If the law plus the true masses of a pair of objects and the distance between them and the assumption that no other forces are at work (plus $f = ma$) entail that the objects should exhibit an acceleration of α and one observes that the acceleration is β (where α and β differ only negligibly), then the idealized law plus the partially specified initial conditions are explanatory.⁸

If (7) and (8) do explain (9), the idea that explanations are arguments appears even more doubtful than Salmon (1984) argued that it is. Salmon's

point is that the *explanans* can confer a low probability on the *explanandum*. However, we don't think that (7) and (8) confer a probability on (9) at all. What is the probability that the observed trait values (the β_i 's) will be close to α , given that α is the trait value that should evolve in an idealized circumstance that does not obtain? We don't know, but it isn't necessary to know this. Propositions (7) and (8) explain (9) even though they do not tell you what the probability of that proposition is. In this type of explanation-by-idealization, conditions (i)–(iii) are satisfied, but (iv) is not.

We began by criticizing Cartwright for drawing an invidious distinction between a conditional and its contrapositive. We then showed how her argument can be reconstructed without requiring that a conditional and its contrapositive apply to different things. This new argument reaches a slightly different conclusion from Cartwright's; we showed how certain sorts of dynamical laws cannot figure in covering-law explanations that are causal. We then tried to flesh out Cartwright's idea that explanation-by-idealization requires a new account of explanation. A causal model contains an idealization when it correctly describes some of the causal factors at work, but falsely assumes that other factors that affect the outcome are absent. The idealizations in a causal model are *harmless* if correcting them wouldn't make much difference in the predicted value of the effect variable. Harmless idealizations can be explanatory, as is shown by the fact that (7) and (8) help explain (9). In this pattern of explanation, the *explanans* is entirely true; it explains the *explanandum*, not by entailing it or by conferring a probability on it (high or low), but by showing that the value described in the *explanandum* is close to the value predicted by the idealization.

ACKNOWLEDGEMENTS

We thank Michael Byrd, Nancy Cartwright, John Earman, Ellery Eells, Berent Enç, Branden Fitelson, Malcolm Forster, Daniel Hausman, and Marc Lange for comments and suggestions.

NOTES

¹ For this issue see Forster (1988a, b), Creary (1981), Chalmers (1993), Earman and Roberts (1999), and Needham (1991). Although all these papers are somewhat relevant to Cartwright's argument above, Forster and Creary specifically address Cartwright's challenge concerning the reality of component forces.

² Although Cartwright says that all laws require *ceteris paribus* modifiers, it is clear from her discussion that in the case of the fundamental laws of physics, she thinks the *ceteris paribus* modifiers are exactly specifiable. This contrasts with the views of philosophers (e.g., Schiffer (1991)) who think that the *ceteris paribus* clauses used in special science generalizations aren't exactly specifiable.

³ Cartwright (1999) says that she still holds the view of laws she defended in Cartwright (1983).

⁴ It is arguable that contraposition is not always valid. Consider the following example: "If I made a mistake, then I didn't make a big mistake". The contrapositive of this conditional is "If I made a big mistake, then I didn't make a mistake". However, contraposition is valid for the laws we will consider. Detailed discussion of this issue can be found in Jackson (1991).

⁵ What if a *backwards* deterministic law contained an idealization in its antecedent? The law will have the form "if *I* holds at time *t*, then *C* holds at time (*t* - Δt)", where *I* involves an idealization. The contrapositive of the law says "if not-*C* holds at (*t* - Δt), then not-*I* holds at *t*". If not-*C* and not-*I* both apply to real systems, then this law can be used in a Hempelian explanation. And if one does not prohibit "negative properties" such as not-*C* from being causes, the contrapositive seems capable of providing a covering-law *causal* explanation. This is why, in the case of deterministic laws, we have limited our argument to *forward*-directed laws that contain idealizations in their antecedents. We owe this point to John Earman.

⁶ Talk of all other evolutionary forces being absent sometimes means that some quantitative variable has a value of zero (e.g., the mutation rate), but at other times it means that certain idealizations are in place (as in the assumption of asexual reproduction). As Cartwright (1983, p. 45) says, it is sometimes apt "...to read '*ceteris paribus*' as 'other things being *right*'".

⁷ This answer does not involve a commitment to adaptationism, which can be thought of here as the view that (7) and (8) are usually true if (9) is; for discussion, see Sober (1993a). Nor does it oblige one to accept the following generalization:

If organisms are fitter the closer they are to the optimal value α and if the forces other than selection are of only negligible value, then the organisms in the population should exhibit trait values close to α .

It is possible that the true but unknown underlying laws exhibit sensitivity to initial conditions.

⁸ Selection and Newtonian gravitation are each construed as deterministic forces within their respective theories. When an idealization concerns a force whose effects are described probabilistically, α should be interpreted as an expected value.

REFERENCES

- Cartwright, N.: 1983, *How the Laws of Physics Lie*, Clarendon Press, Oxford.
 Cartwright, N.: 1999, *The Dappled World: A Study of the Boundaries of Science*, Cambridge University Press, Cambridge.
 Chalmers, A.: 1993, 'So the Laws of Physics Needn't Lie', *Australasian Journal of Philosophy* **71**, 196–207.

- Creary, L. G.: 1981, 'Causal Explanation and the Reality of Component Forces', *Pacific Philosophical Quarterly* **62**, 148–157.
- Earman, J. and Roberts, J.: 1999, 'Ceteris paribus, There are No Provisos', *Synthese* **118**, 439–478.
- Forster, M.: 1988a, 'The Confirmation of Common Component Causes', in A. Fine and J. Leplin (eds), *PSA*, Vol. 1, pp. 3–9.
- Forster, M.: 1988b, 'Unification, Explanation, and the Composition of Causes in Newtonian Mechanics', *Studies in History and Philosophy of Science* **19**, 55–101.
- Hempel, C.: 1965, 'Studies in the Logic of Confirmation', in *Aspects of Scientific Explanation*, Free Press, New York.
- Jackson, F. (ed.): 1991, *Conditionals*, Oxford University Press, Oxford.
- Maynard Smith, J.: 1978, 'Optimization Theory in Evolution', *Journal of Ecology and Systematics* **9**, 31–56. Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*, MIT Press, Cambridge, MA, 1994.
- Needham, P.: 1991, 'Duhem and Cartwright on the Truth of Laws', *Synthese* **89**, 89–109.
- Salmon, W.: 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton, NJ.
- Schiffer, S.: 1991, 'Ceteris paribus Laws', *Mind* **100**, 1–17.
- Sober, E.: 1984, *The Nature of Selection*, The University of Chicago Press, Chicago and London.
- Sober, E.: 1993a, *Philosophy of Biology*, Westview Press, Boulder, CO.
- Sober, E.: 1993b, 'Temporally Oriented Laws', *Synthese* **94**, 171–189. Reprinted in E. Sober, 1994, *From a Biological Point of View*, Cambridge University Press, New York.

University of Wisconsin-Madison
Department of Philosophy
5185 Helen C. White Hall
600 North Park Street
Madison, WI 53706
U.S.A.
E-mails: melgin@students.wisc.edu
ersober@facstaff.wisc.edu