Teleology as Higher-Order Causation: A Situation-Theoretic Account

ROBERT C. KOONS
Department of Philosophy, University of Texas at Austin, U.S.A.

Abstract. Situation theory, as developed by Barwise and his collaborators, is used to demonstrate the possibility of defining teleology (and related notions, like that of proper or biological function) in terms of higher order causation, along the lines suggested by Taylor and Wright. This definition avoids the excessive narrowness that results from trying to define teleology in terms of evolutionary history or the effects of natural selection. By legitimating the concept of teleology, this definition also provides promising new avenues for solving long standing problems in the philosophy of mind, such as the problems of intentionality and mental causation.

Key words: teleology, causation, biological function, proper function, mental causation, situation theory

1. Introduction

In recent years, theories of teleology have come to play a prominent role in theories of cognition, intentionality and the mind. A teleological approach characterizes the work of Dretske, (1988); Stampe, (1977); and others at Wisconsin, (Millikan, 1984; Papineau, 1993; Lycan, 1996). Teleology is especially promising as an explanation of intentionality, since one of the most challenging tasks of an informative account of intentionality is that of explaining the possibility of error or misrepresentation. Ever since the Eleatic philosophers, we have been puzzled by the power of intentionality to relate the mind to non-existent states of affairs. Teleology offers the possibility of grounding the distinction between veridical and erroneous representation in the distinction between function and malfunction, between success and failure at carrying out the mind's natural functions. Moreover, the highly intensional nature of all causal contexts, including teleological contexts, can be used to explain the intensionality (with an 's') of thought. An account that links teleology and causation holds promise for resolving another thorny problem in the philosophy of mind: the problem of mental causation, and the more general problem of the causal efficacy of supervenient properties. Teleology has also been employed with great success by Plantinga (1993), in a recent book on epistemology. Teleology also offers great promise in solving the problem of qualia, and in providing a richer account of personal identity than is offered by mere spatiotemporal continuity.

There is much work to be done on articulating the ontological commitments, semantic underpinnings, and logical principles of teleological thought. The most promising direction to emerge in the literature is that taken by Taylor (1964) and Wright (1976), in which teleology is understood in terms of a kind of higher-order

Minds and Machines **8:** 559–585, 1998.

causation, causal connections in which first-order causal connections act as a causal factor. Making sense of such higher-order causation is no easy matter.

Situation theory (Barwise, 1983, 1989, 1987, 1997; Devlin, 1991) is uniquely suited to providing a set of tools for this urgent task. Situation theory is distinctive in positing a set of partial versions of the world, viz., the situations or situation-tokens. The partiality of situation-tokens is modeled by means of a three-valued semantics (typically, strong Kleene semantics) relating abstract situation-types and concrete situation-tokens. If this partiality is extended to the realm of types involving modality and objective chance, we can use an algebra of situation types and tokens to represent the difference that these modal facts and facts about probabilistic propensities make to the course of events in the world. In other words, we can develop a formal theory of higher-order causation, of just the sort needed to make sense of the Taylor-Wright theory of teleology.

The theory of causation requires two primitive relations between situation tokens, the part-whole relation $\sqsubseteq$ axiomatized by standard mereology, and a relation of causal priority, $\prec$. Causation as a relation between tokens can be resolved into two components: (i) the causal priority of the cause to the effect, and (ii) the existence of a modal or stochastic constraint, making the effect either necessary or highly probable, conditional on the actuality of the cause. The picture according to which a cause necessitates its effect sits nicely within the viewpoint of traditional determinism, while the picture that includes only a probabilistic relation between the cause and effect is commonly associated with an indeterministic view of the world. In this paper, I will present formal versions of both conceptions of causation, although I personally believe that very strong reasons exist for preferring the second. Still, the greater simplicity of the necessitarian model makes its inclusion useful for expository purposes.

First, I will discuss three approaches to the analysis of teleology: the Taylor/Wright account, that employs higher-order causation. Woodfield's account, (Woodfield, 1976) that introduces the idea of the welfare of the organism, and accounts that define teleology in terms of the workings of natural selection. I end up with a version of the Taylor/Wright account, modified in light of Woodfield's criticisms. I then modify the Taylor/Wright definition still further, eliminating the retrospective, historical dimension. According to the new definition, the teleological properties of a system depend only on its present internal organization. and on general facts about causality and objective chance.

In the following section I argue that, although teleology is not best thought of as defined in terms of natural selection, nonetheless, natural selection provides good grounds for believing in real, as opposed to merely apparent, teleology in nature.

In Section 4, I turn to the problem of providing a formal theory of the kind of higher-order causation needed to make sense of the Taylor/Wright definition. I argue that Christopher Hitchcock's approach, (Hitchcock, 1996) which relies on Ellery Eells's definition of causation, (Eells, 1991) is inadequate, and I explain why I think situation theory is needed.

In Sections 5 and 6,1 develop, first, a deterministic model of higher-order causation, and, then, an indeterministic model. These are followed by Section 7, in which I demonstrate that teleological explanation can indeed by modeled successfully by situation theory.

2. Three Definitions of Teleology

The twentieth century has been characterized by an intensifying of efforts at clarifying the logic, semantics and metaphysics of teleology, rectifying the unfortunate neglect of the topic in modern philosophy since Leibniz (1988). One of the most influential and attractive accounts was that of Taylor (1964), in his 1964 *The Explanation of Behavior*. Taylor's influence can be seen in most contemporary accounts, including those of Wright, (1976), Woodfield, (1976) and Millikan, (1984).

I will use Wright, Woodfield and Millikan as paradigms of three competing accounts of the nature of teleological function. These three accounts are the causal, the normative, and the Darwinian, respectively. The Darwinian account has two versions, one retrospective (Millikan) and the other prospective (Bigelow and Pargetter, 1987).

2.1. The taylor/wright account

According to both Taylor (1964) and Wright (1976), a state B occurs for the sake of state G just in case (i) B tends to bring it about that G, and (ii) B occurs because it tends to bring it about that G. This is clearly an instance of higher-order causation: the causal connection between B and G figures in the causation of instances of B. The formal theory of causation that I develop in Sections 5 and 6 is designed specifically to explicate this this sort of possibility.

Millikan (1984) has argued that reliance on this sort of higher-order causation makes sense only if we make explicit reference to the past. She argues that clause (ii) must be replaced by one that reads: (ii') the present token of B occurs because past instances of B tended to bring it about that G.

Wright explicitly rejects this amendment, to which Millikan responds:

> Wright says that the formulation "because X does Z" does *not* reduce to "because things like X have done Z in the past." Rather, we are asked to accept that X might be there now because it is true that *now* X does or X's do result in Z. How the truth of a proposition about the present case can "cause" something else to be the case *at present* is not explained. (Millikan, 1989, p. 299, note 7)

Millikan overlooks two facts. First, the fact that X's tend to bring about Z is not a fact about the present case: it is a timeless, eternal fact about the modal and stochastic structure of the world. Second, Millikan overlooks the fact that such eternal facts can enter into causal explanations of present conditions, as I will argue in detail in Section 7.

We can distinguish a number of interesting varieties of teleological connection. First of all, we can distinguish between intrinsic and extrinsic purpose. For example, the bird of a wing exists for the sake of flying, and this is a case of intrinsic purpose. In contrast, seeds serve the purpose of feeding the bird: a case of extrinsic purpose. Another distinction we can make is that between productive and informational functions. The Taylor/Wright definition specifies one important class of functions: the productive functions. However, there are also receptive or informational functions. For example, the eye has the function of registering the existence of certain kinds of objects in the environment. This function is not a matter of the eye's effect on the environment, but the reverse: the environment's effect on the eye. In earlier work, I defined a relation of information (or potential information), building on the ideas of Dretske, (1991), Koons, (1996). We can say that a particular pattern of retinal stimulation ϕ has the intrinsic function in s (relative to v) of carrying the information that ϕ just in case the pattern ϕ exists because it carries (in organisms of type v) the information ψ. We might say that when a state occurs that has the function for an organism to carry potential information of a certain kind, then that information has become actual for that organism.

2.2. From Woodfield and Bedau to Aristotle

Woodfield (1976) argues that the Taylor/Wright account gives a necessary, but not a sufficient, condition for teleo-functionality. He urges that we must add a normative element, requiring that the functional state contribute to the well-being of the organism. An example created by Plantinga (1993), gives some support to Woodfield's contention. We are to imagine a world in which a Nazi-like regime institutes a dysgenics program, aimed at a hated minority race. A harmful mutation is introduced into the minority population, that renders the bearer nearly blind, and makes attempted seeing painful. The Nazi breeders gradually eliminate all of the members of the minority race without the gene, by testing for signs of faulty and painful vision. In such a case, the defective gene appears to satisfy Wright's criterion, since part of the causal explanation of the presence of the gene in the population is the deleterious effect of the gene on the bearer's vision. Yet, it would seem odd, at the very least, to say that the gene had the function (and not just the effect) of impairing vision.

There are a number of other examples that also suggest that the Wright definition is too broad. Any stable feature of the inanimate world characterized by feedback loops, that is, any genuine case of dynamic equilibrium, will be describable as instantiating teleofunctionality, according to Wright's definition. Suppose, for example, that the presence of ice in a rock crevice causes the crevice to remain open.[1] In this case, the existence of ice in the crevice is caused by the power of the ice to keep the crevice open. The ice has the Wrightian function of keeping the crevice open. Similarly, if the rapid flow of water in a channel keeps the channel from silting up, we would have to say that the water flow had the function of preventing the deposition of silt, since in the absence of that causal connection, the silt would prevent the water from flowing so rapidly. In these cases, Woodfield would argue, there is no genuine teleofunction, since ice deposits and water flows have no welfare.

If we merely add the condition of welfare-enhancement to Wright's definition, however, we would seem to have only a verbal difference, one definition for Wright-functions, and another for Woodfield-functions. with the dispute concerning only the appropriate meaning for the English word 'function'. It is possible, however, to reconstrue Woodfield's position as an alternative metaphysical account. We could take Woodfield as claiming that there is a metaphysically distinguished class of Wright-functions: those which exist because they contribute to the welfare of their bearers. Such an account gives a real causal role to the property of goodness (goodness for some kind of organism), resulting in something very close to Plato's theory of the Good, (Plato, 1986, Book VI).

Bedau has also argued that an evaluative element is essential to teleology (Bedau, 1992). Bedau distinguishes "three grades of evaluative involvement". In the first grade of involvement, we define the proper function of ϕ to be ψ by requiring that ϕ brings about ψ, and ψ is good. This adds goodness to a pre-Wrightian, dispositional account of function. In the second grade, we incorporate Wright's definition and add that ψ is good as an additional and separate condition. That is, we require that the thing has ϕ because ϕ brings about ψ, and, in addition, that ψ is good. Finally, in the third grade, we make include the goodness of ψ within the causal explanation of ϕ: the thing has ϕ because both ϕ brings about ψ and ψ is good.

Let $\gamma(v)$ represent the situation-type in which the welfare of the type of organism whose time-slices are of type v. We could then define a third-grade or Platonic function (relative to kind v) as one in which the end promoted also promotes $\gamma(v)$, and the fact that it does so is also causally relevant to the existence of the functional state. This additional condition, which we can call the 'Platonic condition', requires that there be a causal connection between ψ (the Wright-functional end of ϕ) and the welfare of the organism (qua member of the background kind v).

There is an alternative, somewhat more deflationary account, of the role of goodness in a Bedavian third-grade definition of teleology. A thing is capable of well-being just in case the sum of its Wright-functions forms a highly coherent, mutually supportive totality. A Wright function counts as a genuine teleofunction just in case it coheres in this sense with the well-being of its possessor. This sort of an account also has echoes of Platonic themes, in this case the close connection for Plato between well-being and *harmony* (Plato, 1986, Book IX). A thing, like an organism, with a largely harmonious set of Wright-functions is capable of well-being; inanimate objects, with largely unrelated, discordant Wright-functions, are not. Plantinga's example of the dysgenic gene can be excluded, since, although the gene does have a Wright function, this function does not cohere well with the rest of the Wright functions of its human hosts.

There is one more refinement that needs to be made, bringing this deflationary account closer to the Platonic one. We need to distinguish between those cases in which the Wright-functions of a thing are harmonious, but the harmony of the functions is merely coincidental, and those cases in which the harmony of the Wright-functions is itself functional, contributing, perhaps, to the adaptive fitness of the organism. According to the deflationary account, harmony is constitutive of the good. Hence, both cases are cases of organisms with a standard of well-being. Alternatively, we might insist that the harmony of Wright-functions must itself be *explained* by reference to the good. This moderate position we might call an "Aristotelian" theory of the good. According to this account, we can define the good of thing in the following way:

Aristotelian Definition of the Good

– A thing has a good if and only if it has proper functions.
– The good of a thing consists in the successful exercise of its primary proper functions.

Aristotelian Definition of Proper Function. A state ϕ has the proper function ψ, in kind υ if and only if:

1. The fact that things in kind υ have state ϕ is causally explained (at least in part) by the existence of a causal law linking ϕ & υ to ψ, as cause to effect (Wright's condition).
2. The system of functions $< \phi_i, \psi_i >$ meeting condition 1 for υ forms a mostly harmonious, mutually supportive whole, and the $< \phi, \psi >$ function contributes to this harmony.
3. The existence of things of kind υ is causally explained (at least in part) by the harmony mentioned in condition 2.

This Aristotelian definition is stronger than the deflationary account. since it requires more than the bare fact of the existence of a harmony among Wright-functions. At the same time, it takes on much less ontological burden than the full-blown Platonic account, since it does not have to postulate goodness as a primitive causal factor that explains the existence of Wright-functions. Its combination of sober realism with ontological moderation seems to justify calling it "Aristotelian", at least in inspiration.

Bedau argues that biology makes use only of first and second-grade functions. He denies that third-grade functions have a legitimate place in the modern, scientific picture of the world. However, he reaches this conclusion because he overlooks the possibility of an Aristotelian version of third-grade evaluative involvement. In fact, it is the third grade, understood in this deflationary way, that is needed to distinguish the functionality of organisms and artifacts from self-perpetuating equilibria in the inanimate world.

For an organism to have a harmonious set of functions, it is not necessary that it have no dysfunctional features, nor do we need to exclude the existence of a moderate degree of competition and interference between the organism's various functions. Let us say that function x harmonizes with system S just in case, for *many*, but not necessarily all, members y of S, the fulfillment of x increases the probability of the fulfillment of y, and, for *most* but not necessarily all members y of S, the fulfillment of x does not significantly decrease the probability of the fulfillment of y. A system of functions S is harmonious if nearly every member x of S harmonizes with $S - \{x\}$.

This definition of harmony is not entirely successful, however, because it does not take into account the existence of secondary and tertiary functions. For example, the body may respond functionally to a condition in which it has suffered massive injuries by radically lowering the metabolic rate. This functional response is fulfilled only when many other functions have failed; hence, the fulfillment of this secondary function significantly lowers the probability of the fulfillment of most of the body's functions, since it entails that these functions have in fact failed. It is possible that an organism could exist most of whose functions were secondary ones. In response, let us say that a function x compensates for a set of functions T just in case the successful fulfillment of x entails that none of the members of T are fulfilled and is causally posterior to the failures of the members of T. A function x meta-assists y relative to T just in case x compensates for T and the fulfillment of x increases the probability of the fulfillment of y, conditional on the failure of the members of T. We can then weaken the definition of harmonizing with a system by requiring only that the function meta-assist some of the members of the system, relative to some proper subset of the system. A system is harmonious if most of its members harmonize (in the new, weaker sense) with the remainder of the system, and many of its members harmonize (in the first, stronger sense) with it.

An organism fighting off an infection, or infested with a parasite. is the locus of two disjoint systems (its own and the parasite's), each internally harmonious, and each in conflict with the other. In cases of symbiosis, we can identify two disjoint systems, even though they are mutually supportive, since the ancillary connections between the two systems are much fewer and weaker than those within each one. Cases such as that of the mitochondria lie on the vague boundary between organic unity and close and long-established symbiosis.

Any organism will suffer from a certain degree of dysfunctionality. The standard is one of substantial harmony among functions, not ideal or optimal harmony. The function of x is not determined by working out what x is optimally designed for, but by working out whether the most likely explanation for the origin of x involves a causal connection between x and some effect. For example, there are cases of selfish DNA, genes that take control of the gene replication process, producing multiple copies of themselves on the chromosome, despite the fact that they interfere with the organism's fitness. These selfish genes constitute a kind of self-perpetuating genetic illness, a chromosomal parasite. The existence of such imperfections in the chromosomal system does not pose any challenge to the obvious fact that the function of the system includes cell reproduction and protein synthesis.

For the purposes undertaken in this paper, I will use the Aristotelian definitions of good and of proper function as my working hypotheses. I believe that the Aristotelian definition is weak enough to include as proper functions everything we would want to attribute as such to organisms and to artifacts, while excluding any property in the inanimate, natural world as functional.

2.3. NATURAL-SELECTION ACCOUNTS

Very roughly, Millikan (1984) defines the relation of functionality in terms of actual contribution to the survival and reproduction of the organism's ancestors. The eye has the function of registering information of a certain kind because the fact that similar organs in the ancestors of the organism in question contributed to the successful reproduction of those ancestors by registering such information. Millikan's account is explicitly retrospective, which invites certain kinds of objections. The first appearance of a new adaptation is always non-functional, since it cannot acquire a function until it has actually contributed causally to successful reproduction. This applies even to artifacts: if I design a widget to perform a task, and it does so and in the very way that I envisaged, it still does not have that function until its success at meeting the need for such functionality results (say, through the marketplace) in the reproduction of duplicate widgets. In addition, on Millikan's account, once a function has been acquired, it can never be lost. The sightless eyes of cave fish still have the function of seeing, and words of contemporary English still carry the meanings of their Indo-European roots. These results seem counterintuitive.

One solution would be to make Millikan's account prospective instead, as Bigelow and Pargetter (1987) have done. On their account, a state has a particular function if the fact that it tends to produce this result enhances the reproductive fitness, here and now, of the organism in question. It is not clear that this strategy will work, however, since it is unclear what "reproduction" can mean in a purely prospective sense. Millikan has the advantage of being able to make reference to an already existing family of similar, self perpetuating structures. Since everything is similar to everything else in some way, it is unclear what "the reproduction of *x*" can mean, in the absence of some already existing class of organisms to which *x* belongs.

There is, however, a more fundamental problem with all of these accounts: the fact that they make the truth of Darwinism a matter of ontological necessity. Surely it is possible, in some suitably broad sense that functional organisms come into existence in the way described in the book of Genesis, even if this is not the way things happened in the actual world. Moreover, it would seem to be possible for there to exist what Sorabji (1964) calls "luxury functions": functions that do not in fact enhance the reproductive fitness of their bearer, and that did not enhance the reproduction of its ancestors. For example, the capacity to appreciate beauty for its own sake, or the ability to track the truth in metaphysical domains, may be genuine functions of the human mind that have nothing to do with reproductive fitness. It is at least possible that such functions exist: our fundamental account of the nature of function should not exclude these possibilities.

It is far more plausible to take natural selection as a mode of explaining how it is that functions exist in the world, not as an account of what it is for something to be a function.

Neander (1991), has defended a natural-selection account of teleology as an analysis of the concept of function, as it figures in the thinking of contemporary biologists. According to Neander, in the specialist language of contemporary biologists, the word 'function' just means 'selected for by nature'. If contemporary biologists have made the truth of Darwinism a matter of stipulative definition, so that to deny the neo-Darwinian synthesis, one would have to deny that biological functions exist, then this would constitute an unjustifiable form of dogmatism, setting up a conceptual barrier to any future theory that might prove superior to the contemporary synthesis. This stipulation would make rational dialogue between Darwinists and contemporary or future critics impossible, since supplanting the present theory would require a conceptual and linguistic revolution.

Moreover, the notions of 'function' and 'natural purpose' have roles to play far beyond the narrow world of biological specialists. Functionality is an important concept in our commonsense view of the world, and it is needed (I will argue) in an adequate theory of epistemology and ethics. The content of such a widely used concept cannot be settled by the linguistic conventions of a specialized community.

2.4. Retrospective versus non-retro accounts

So far, all of the definitions of teleology we have considered have been retrospective in nature, in the sense that the function of a thing depends upon what was involved in causing certain features of that very thing. This would mean that teleo-functions do not supervene on the internal organization of a thing. Two internally indistinguishable systems could have different functions, due to differences in the causal histories involved. For example, a swamp-bird, that forms spontaneously, without evolutionary history, has swamp-wings that, unlike birds' wings, do not have the function of enabling flight, even if the swamp-bird does soar about with apparent facility.

Many philosophers, including Dretske and Millikan, are content to bite the bullet of this consequence. My inclinations are to try to dodge it. Moreover, there is another problem with retrospective natural-selection accounts of teleology. Ironically, they propose an essentially neo-Lamarckian conception of function. According to Lamarckian theory, use must always precede function. It is only after a particular structure or behavior has proved its usefulness in practice that it can be incorporated into the set of adaptations of the individual or population. In contrast, neo-Darwinian theory opens the door to the possibility that a function can emerge spontaneously, by fortuitous mutation. Natural selection explains, not the origin or nature of the function, but its successful perpetuation. This issue is particularly acute when attempting to understand systems of interdependent functions. Consider, for example, the mutually presupposing functions of sexual reproduction. The function of the sperm is to fertilize the ovum; and the function of the ovum is to receive the sperm. Neither can operate before the other is functional. Hence, it is incoherent to insist that the gametes cannot be functional until *past instances* of each have successfully been used in reproduction.

Indeed, in this case, the difficulty for the natural-selection account of functionality is especially acute, since there can be no such thing as gametes *before* the functional system of sexual reproduction has been established. Hence, the functionality of gametes cannot be explained in terms of the previous history of gametes, since there could not, by the very nature of the case, be such a thing.

It is possible to define the function of a thing without building in any conditions about the actual causal history of that very thing. Let us say that the Aristotelian definition of function given above is the definition of 'etiological-function'. Then, we can say that some feature A of some thing x of kind K has function F just in case the objective probability is greater than one-half that something with the internal organization specified by K would have been caused in such a way as to make F the etiological- function of A. For example, the swamp-bird belongs, by virtue of its internal organization, to a class of things B of such a kind that the objective probability is greater than one-half that an arbitrary member of B came into existence through the kind of natural selection responsible for the existence of ordinary birds. Therefore, even though the swamp-bird came about in a very unusual way, a way in which the causal powers of its wing-like appendages had no role, we can still say that the function of these appendages is to enable the swamp-bird to fly.

In contrast, if natural processes accidentally produce something internally indistinguishable from a very crude arrowhead, we do not have to say that its function is to act as the point of an arrow, since the objective probability of the accidental production of such a system is non-negligible. The difference between the swamp-bird and the arrowhead-like stone lies in the astronomical difference in the objective probabilities of the spontaneous generation of each.

In the case of systems of interdependent functions, such as those of the sperm and the ovum, each individual gamete, even the very originals, are such that it is very likely that something so organized resulted from a process that included successful reproduction (i.e., favorable natural selection). Even though the original gametes had no such selective history, it is far more likely (in terms of objective chance) that something so organized is one of the many successful descendants of the original mutants than that it is a product of favorable mutation. Thus, the original gametes were fully functional, despite the fact that their actual history included nothing that satisfies Wright's higher-order condition.

3. Does Darwinism Support Real or Only Apparent Functionality?

Darwin's theory of natural selection (Darwin, 1859) has been taken in two quite opposing ways on the question of its bearing on teleology. American biologist Asa Gray took Darwin's theory as vindicating the reality of biological teleology, and, in a letter to Gray, Darwin himself seems to endorse this inference (Gilson, 1984). In contrast, many philosophers and scientists, including, most recently, Dawkins (1987) and Dennett (1995), have taken the upshot of Darwin's theory to be that all biological functionality is merely apparent, with natural selection explaining the existence, not of real teleology, but only of its appearance in nature.

These two conclusions are most probably based on two different understandings of the nature of teleology. It would seem that those taking the Dawkins/Dennett line assume that the existence of a function entails the existence of a designer or creator, whose prior intentions, or whose intentions plus their effective realization, constitute the functional character of the product. Plantinga (1993) in his recent book, *Warrant and Proper Function*, explicitly affirms the existence of this implication. I have two reasons for demurring. First, it seems that something like the accounts of Wright or Woodfield are adequate characterizations of functionality, with the products of intentional design clearly falling under the definiens, without necessarily exhausting its extension. Second, I hope to give an account of intentionality in terms of teleo-functionality (roughly, a state represents a fact just in case it has the function of carrying the corresponding potential information), so accepting Plantinga's analysis would doom such an analysis to vicious circularity.

Consider again the case of the bird's wing's having the function of enabling flight. The causal connection between the presence of wings and flight was itself a higher-order cause of the successful survival of winged ancestors of existing birds. A given stage of a winged-bird organism is caused to be bird-stage, and hence is caused to be winged, by these earlier success in survival. Thus, there is an indirect causal connection between, on the one hand, the causal wings-to-flight connection, and, on the other hand, the presence of wings in the given specimen. Wright's definition is satisfied. Moreover, wingedness is part of a system of harmonious functions in the form of life of the bird, and the harmony of these functions is itself adaptive. Thus, the more stringent Aristotelian definition is also satisfied.

The connection via natural selection is indirect and retrospective. If all actual teleology were explicable by natural selection alone, we would have to deny the existence of real (as opposed to merely apparent) "luxury functions". However, this would be a consequence, not of an ontological theory (as in Millikan's case), but of biological theory.

Functions are explained by natural selection in an indirect and retrospective manner, but so are functions that are explained by the intentions of a designer. The intentions of the designer mediate between, on the one hand, the causal connection between the trait and its effect, and, on the other hand, the existence of the functional trait in the product, just as the evolutionary history of an organism mediates between these two in the case of natural selection.

4. The Problem with Higher-Order Causation

In a recent paper, Hitchcock (1996) uses Eells's definition of causal relevance (Eells, 1991) to defend the intelligibility of higher-order causation. According to Eells, a property ϕ is (positively) causally relevant to property ψ in population p just in case the objective probability $Pr(\psi/\phi\&\eta)$ is strictly greater than $Pr(\psi/\neg\phi\&\eta)$, for all homogeneous background contexts η. In applying Eells's definition to higher-order causation of the kind employed in the definition of teleology, we must suppose that ϕ is itself a property involving causal relevance. For example, Wright's definition of ϕ's having ψ, as a function would, when translated into Eells's definition of causation, come out as something like this (ignoring the background contexts for the sake of simplicity):

$$Pr(\phi/[Pr(\psi/\phi) > Pr(\psi/\neg\phi)]) > Pr(\phi/[Pr(\psi/\phi) \leq Pr(\psi/\neg\phi)])$$

This account depends on making sense of higher-order objective chance, in particular, of making sense of the present objective chance of ψ, given ϕ, and of ψ, given $\neg\phi$ being other than they actually are. As Hitchcock notes, it is very hard to see how to make sense of the present objective chance of any present objective chance being either 1 or 0. In the present state of the world, whatever factors that determine objective chance are either definitely present or definitely absent, so the actual objective chance of any proposition is fully determined.

Hitchcock attempts to circumvent these problems without resorting to situation theory by introducing the parameter of *populations*. He suggests that we treat the objective chance of ψ given ϕ, and of ψ, given $\neg\phi$ as properties of various actual and hypothetical populations. The claim about higher-order causation is then taken to be a claim about a super-population, whose individual members are actual or hypothetical populations. However, Hitchcock has merely sidestepped the problem. To make sense of this solution, we must know two things: (i) which hypothetical populations to include as members of the superpopulation, and (ii) what probability measure over these hypothetical populations to use in computing the higher-order probability. To have a principled solution to these two problems, we would have to know the objective chance of the various objective chances represented in the hypothetical population. However, it was exactly the unavailability of such higher-order objective chances within the conventional possible-worlds approach that led to the impasse described above.

This collapse of higher-order objective chance to triviality could be avoided if we consider partial worlds or situations, this conclusion no longer holds. A situation is partial, so many of the factors that determine objective chance are undetermined in a given situation. We can then, sensibly talk about a hierarchy or cascade of objective chances. Meaningful higher-order objective chance could exist, whenever there are well-defined objective chances of certain factors, whose presence or absence would, in turn, determine the objective chance of other factors.

However, there is, as we shall see, another approach within situation theory to defining the causal relevance of facts about objective chance, an approach that does not depend on making sense of higher-order objective chance. Rather than asking how the objective chance of ϕ depends on the objective chance of ψ, given ϕ, or of ψ given $\neg\phi$, we can instead ask whether "deleting" facts about the causal connections between ϕ, and ψ, from particular situations leaves enough facts behind to enable those situations to cause the relevant instances of ϕ. This talk about "deleting" facts from situation-tokens is metaphorical. We start with a token that supports the causal connection between ϕ and ψ, and then we consider proper parts of this token that do not support this connection and ask of these parts whether they support enough facts to enable them to count as causes of the instances of ϕ in question. In this case, it is the the *indispensability* of facts about causal connections as incorporated in *parts* of actual situation, rather than the probabilistic relevance of those facts to abstract properties, that determines the existence of a causal connection.

In the following two sections, I will lay out two theories of causation, one deterministic and one indeterministic, using situation theory. In each of these, the problem of determining the causal relevance of causal connections will prove tractable, and we will not be forced to introduce anything as recherche as higher-order objective chance.

5. Higher-order Causation: The Deterministic Conception

5.1. BASIC ONTOLOGY

In this section, I will introduce the model structures to be taken as formal representations of real possibilities concerning causation. These structures incorporate two kinds of individuals: situation-tokens and situation-types. (Barwise, 1983, 1987, 1989). Actual situation-tokens are to be thought of as real, concrete parts of the world, analogous to Davidsonian events (Davidson, 1980). Merely possible situation-tokens are abstract objects, constructible from actual tokens and types, representing possible but unrealized actualities. Each token carries a certain amount of information or fact about the world: these units of fact are represented as situation-types.

5.1.1. Classification Systems

A classification system consists of a set of tokens, a set of types, and a binary relation on the two sets (the classification relation).[2] For my purposes, the set of tokens will be a set of situation-tokens, the set of types situation-types, and the classification relation $\models$.

I will assume that the set of types is closed under the Boolean operators '$\neg$' and '$\vee$'. The classification relation $\models$ can be constrained to satisfy such logical principles as De Morgan's laws, distribution, associativity and commutativity of $\vee$, and the laws of double negation. In addition, if $s \models \phi$, then $s \not\models \neg\phi$, but not necessarily vice versa. (So, '$\neg$' on types can represent weak or internal negation.)

5.1.2. Models

In the most general case, a model would contain a set of tokens, *Sit*, together with a function $\mathcal{S}$ that assigns a classification system to each situation. In addition, we need two partial orderings on situation-tokens, $\sqsubseteq$ and $\prec$. The first represents the part-whole relationship of standard mereology. The second, a strict partial well-ordering, represents the relation of causal precedence. In representing causation, we can look at a simpler, special case, one in which all of the classification systems share the same set *Typ* of types and the same classification relation $\models$. In this special case, the function $\mathcal{S}$ assigns to each situation-token s a subset of *Sit*. The tokens in $\mathcal{S}(s)$ are those that are possible alternatives, from the perspective of s, that is, the set $\mathcal{S}(s)$ represents the modal facts about the world as they are supported by s.

$\mathcal{W}(s)$ shall be the $\sqsubseteq$-maximal members of $\mathcal{S}(s)$. These situations represent possible "worlds", from the perspective of s.

Consequently, a *standard, deterministic model* $\mathcal{M}$ consists of an n-tuple, $< Sit, Typ, \models, S, \sqsubseteq, \prec >$, where:

- *Sit* is a nonempty set, the set of situation-tokens.
- *Typ* is a nonempty set of situation-types, closed under the Boolean operators $\vee$ and $\neg$.
- $\models$ is a binary relation on $Sit \times Typ$.
- $\mathcal{S}$ is a function from *Sit* to the powerset of *Sit*.
- $\models$ is a partial ordering of *Sit*. There is a set $\mathcal{W}$ of maximal situations (worlds). Every situation is extended by some world.
- $\prec$ is a partial ordering of *Sit*.

5.1.3. Persistence, Exclusion and Saturation

One situation excludes another whenever there exists no situation containing both of them as parts. We can abbreviate this relation as $s \perp s'$. If we assume that all situations are coherent, in the sense that there is no type ϕ such that the situation belongs to both ϕ and $\neg\phi$, then facts about what situations exclude other situations will be constrained by facts about the persistence of types, that is about when a whole inherits the types of its parts.

There are four forms of persistence that seem plausible:

1. *Global persistence*. If a part belongs to the type, so does the whole.
2. *Synchronic persistence*. If s belongs to the type, $s \sqsubseteq s'$ and no part of either is causally prior to any part of the other, then s' also belongs to the type.

3. *Punctual persistence.* If s belongs to the type, $s \sqsubseteq s'$, and exactly the same situations are prior to both, then s' also belongs to the type.
4. *Non-persistence*. There is no condition that guarantees that when a part belongs to the type, so does the whole.

A globally persistent type represents an eternal fact (such as a modal or mathematical fact), or a fact that includes reference to particular individuals (or places) at a particular time (such as ‘Clinton was speaking at noon, July 3, 1994’). A synchronically persistent type represents a fact in which individuals or places involved are specified, but not the time, such as ‘Clinton speaking at the White House’. A punctually persistent type represents a purely qualitative type, one in which no particular individual, place or time is specified, such as ‘a man speaking on a platform’. In the case of a punctually persistent type, the spatio-temporal location of the fact is fixed by the causal antecedents of the situation-token belonging to the type.

I will assume that all types are at least punctually persistent.

5.1.4. Identity conditions for tokens

I will assume that each token has three kinds of properties essentially: its types (representing its intrinsic character or quality), its parts, and the network of its causal antecedents (representing its backward time-cone). The third assumption is a generalization of the Kripkean intuition (Kripke, 1972) that the origin of a thing is always essential to it. It seems plausible to suppose that a particular event could not have been the very event it is if either the intrinsic character of the event were different, or if the causal chain leading up to the event were different. In contrast, the subsequent course of events, causally posterior to an event, are not essential to its identity. The very same event could exist in different worlds, with different subsequent histories.

If we make these assumptions, then any possible token could be represented as an ordered pair, consisting of a set of coherent types, and a causal tree of possible tokens, rooted in the immediate causal antecedents of the token. I would not want to *identify* a real situation token with such an ordered pair. There is, however, a homomorphism from actual tokens to the set of such pairs. Pairs that do not represent actual situation-tokens can be taken as representing merely possible tokens.

5.1.5. The Causal Priority Relation

The causal priority relation $\prec$ cannot be identified simply with causation. Instead, it represents a necessary pre-condition for causation. In fact, under the assumption of determinism, these three causal notions are interdefinable:

– $\prec$, the relation of causal priority.
– $\triangleright$, the relation of being a total cause of.

– $\rightsquigarrow$, the relation of being an essential part of a total cause of, Mackie's INUS condition: "an Insufficient but Necessary part of an Unnecessary but Sufficient condition" (Mackie, 1965).

In addition, we can say that two tokens are *coincident* if they share exactly the same causal antecedents.

Definition 5.1. (Coincidence). $s \approx s' \leftrightarrow \forall s'' \in Sit(s'' \prec s \leftrightarrow s'' \prec s')$

I will assume that the causal priority relation is transitive, irreflexive, and well-founded. A token s is immediately prior to s' just in case s is prior to s', and there are no intermediate tokens.

$$(s \prec_0 s') \leftrightarrow_{df} (s \prec s')\&\neg\exists s''(s \prec s''\&s'' \prec s')$$

The three causal relations $\prec, \rightsquigarrow$ and $\rhd$ are such that any two of these can be defined in terms of the third, together with the mereological part-whole relation $\sqsubseteq$:

– $s' \prec s'$ iff there is a w such that s is a part of the mereological sum of INUS causes of s' in w (or equivalently, iff s is a part of the mereological sum of minimal total causes of s' in w).
– $w \models s \rightsquigarrow s'$ iff s is part of some minimal total cause of s' in w.
– $w \models s \rhd s'$ iff $s \prec s'$, $s, s' \sqsubseteq w$, and s necessitates s'.

More or less arbitrarily, I will take the causal priority relation to be primitive and define the other two in terms of it. In this case, the first condition above imposes a minimality requirement on the extension of $\prec$ in a model.

5.2. Constraints and Causation at the Token Level

On the deterministic conception, a token s causes token s' if three conditions are met: (i) token s is actual, (ii) s is causally prior to s', and (iii) the actuality of s constrains the world to contain s' as well, in other words, s objectively necessitates s'. This notion of constraint or necessitation can be defined thus.

Definition 5.2 (Token-to-token Constraint). Token s supports the constraint $(s_1 \vdash s_2)$ holds between tokens s_1 and s_2 if and only if: every token s' that represents a possible world or maximal token from the perspective of s and that does not exclude the possibility of s_1 includes s_2 as a part.

We can represent this definition (and its negation) symbolically as:

$$s \models (s_1 \vdash s_2) \Leftrightarrow$$
$$\forall s' \in \mathcal{W}(s)(\neg(s_1 \perp s') \rightarrow s_2 \sqsubseteq s')$$
$$s \models \neg(s_1 \vdash s_2) \Leftrightarrow$$
$$\exists s' \in \mathcal{W}(s)(s_1 \sqsubseteq s'\&s_2 \perp s')$$

This token-to-token constraint relation is one of strict necessitation: every world containing the first situation must also contain the second.

We can now turn to the case of *causal* constraints. We are presupposing a deterministic model, which in the case of tokens consists of two theses: causes must be actual, and causes constrain their effects to be actual as well. Given the definition of causal constraint, it will follow that if s is a possible world, and $s \models (s_1 \rhd s_2)$, then both s_1 and s_2 must be actual in (i.e., parts of) s.

Definition 5.3. (Token Causation under Determinism). A token s supports a causal connection $(s_1 \rhd s_2)$ if and only if s_1 is part of s, s_1 is causally prior to s_2, and s_1 itself supports the token-to-token constraint $(s_1 \vdash s_2)$.

Notice that, in order for $s_1 \rhd s_2$ to be true in situation s, it is necessary that s_1, and not just s, support the constraint $s_1 \vdash s_2$. This requires that a sufficient number of relevant modal facts are incorporated into the total cause. It is this feature of situation theory that makes possible the representation of the kind of higher-order cansation needed in the definition of teleology. The definition can be represented symbolically as:

$$
\begin{aligned}
&s \models (s_1 \rhd s_2) \Longleftrightarrow \\
&\quad s_1 \sqsubseteq s \& s_1 \prec s_2 \& \mathcal{M}, s_1 \models (s_1 \vdash s_2) \\
&s \models \neg(s_1 \rhd s_2) \Longleftrightarrow \\
&\quad (s_1 \perp s) \vee \neg(s_1 \prec s_2) \vee \mathcal{M}, s_1 \models \neg(s_1 \vdash s_2)
\end{aligned}
$$

5.3. EXTENDED SITUATION-TYPES

Teleology is a higher-order causal connection between situation-types. Consequently, I must introduce definitions of static and causal constraints at the level of types as well as tokens. The formula $(\phi \vdash \psi)$ shall represent the existence of a static constraint involving types ϕ and ψ, and $(\phi| \sim \psi)$ shall represent a dynamic constraint. In order to represent higher-order constraints with full generality, we must treat constraints as constituting situation-types themselves. When a constraint between ϕ and ψ, is supported by a situation-token s, we shall say that s is of a modal type. Hence, we must define a class of *extended situation-types*.

– If ϕ is a basic type and s is a situation token, then $(s : \phi)$ is an e-type.
– If s and s' are situation-tokens, then the following are e-types: $(s \sqsubseteq s')$ and $\Diamond s$.
– If ϕ and ψ, are e-types, so are $\neg\phi$, $(\phi \vee \psi)$, $(\phi \vdash \psi)$ and $(\phi| \sim \psi)$.

The one type that I have not yet mentioned is that of $\Diamond s$. $\Diamond s$ is true at a token s' just in case s is part of one of the members of $\mathcal{W}(s')$, the possible worlds from the perspective of s'.

5.4. CONSTRAINTS AND CAUSATION AT THE TYPE LEVEL

We can now define constraints at the level of situation-types. There are two kinds of type-level constraints to consider: static, and causal (or dynamic). A static constraint of the form $\phi \vdash \psi$ means that any token of type ϕ can be extended to a coincident token of type ψ, i.e., there is a ψ-type *fulfillment* in every world of every token of type ψ. This extension to a coincident token keeps us at exactly the same point in the network of causal connections, hence the epithet “static”. In contrast, a causal constraint moves us from one token to an immediately successive token in the causal order.

Both static and dynamic constraints will be defined in two steps. First, I will define extensional, local versions of the two constraints, by quantifying over the parts of a single world. Then, I can define modal, intensional versions, by quantifying over the worlds (as defined from the perspective of a given situation).

Definition 5.4. (Static, extensional constraint on types). A situation-token s supports a static constraint $(\phi \hookrightarrow \psi)$, between situation types ϕ, and ψ, just in case, for every situation s' such that $s' \sqsubseteq s$ and $s' \models \phi$, there is a situation s'' such that $s'' \sqsubseteq s$, $s' \approx s''$, and $s'' \models \psi$.

Definition 5.5. (Causal, extensional constraint on types). A situation-token s supports a causal constraint $(\phi \Rightarrow \psi)$, between situation types ϕ, and ψ, just in case, for every situation s' such that $s' \sqsubseteq s$ and $s' \models \phi$, there is a situation s'' such that $s'' \sqsubseteq s$, $s' \prec_0 s''$, and $s'' \models \psi$.

Definition 5.6. (Static, modal constraint on types). A situation s supports a static modal constraint $(\phi \vdash \psi)$ just in case every $w \in \mathcal{W}(s)$ such that w supports $(\phi \hookrightarrow \psi)$.

The standard notions of modality, such as necessity, can be defined in terms of static constraints: $\Box\phi =_{df} (\neg\phi \vdash \phi)$.

Definition 5. 7 (Causal, modal constraint on types). A situation-token s supports a causal constraint $(\phi| \sim \psi)$ just in case every $w \in \mathcal{w}(s)$ is such that w supports $(\phi \Rightarrow \psi)$.

Modal type constraints, both static and dynamic, give rise to a distinctive form of modal logic. Since we are working with partial, three-valued worlds, substitution into modal contexts is permissible only if the relevant types are strong-Kleene equivalent, not just classically equivalent. For example, ϕ, and $((\phi\&\psi) \vee (\phi\&\neg\psi))$ are classically, but not strong-Kleene, equivalent. This hyperintensionality of causal contexts is vital to their use in explicating teleological and representational properties.

Now that we have constraints on both the level of tokens and that of types, we can speculate about the relationship between the two. Since Hume, many philosophers have taken the view that token causation never occurs in the absence of a type-level constraint. This amounts to the rejection of what is known as ‘singular causation’. Hume’s proposal can be expressed as a requirement of the supervenience of token constraints on type constraints.

Proposition 1 (Hume's Hypothesis). If situation s_1 *supports the token causal constraint* $(s| \sim s')$*, and* $s' \models \psi$*, then there exists a type* ϕ*, such that* $s_1 \models (\phi| \sim \psi)\&(s : \phi)$.

5.5. CAUSAL RELEVANCE

A key notion in the Taylor/Wright definition of teleology is that of causal relevance. Causal relevance is a relation, not between tokens, but between token/type pairs: it is token s, *qua* type ϕ, that is or is not relevant to s', qua type ψ. In defining causal relevance, I will make use of the relation of INUS causation, $\rightsquigarrow$. When a token s is an INUS cause of s', then every part of s is causally relevant to some part of s'. If, in addition, s' is a minimal token for which s is an INUS cause, then every part of s is relevant to every part of s'. In such circumstances, if s is of type ϕ, and s' is of type ψ, it seems reasonable to say that s's being ϕ is causally relevant to s''s being ψ.

Definition 5.8. (Causal Relevance). $(s : \phi) \rightsquigarrow (s' : \phi)$ if and only if (i) $(s \rightsquigarrow s')$, (ii) $s \models \phi$, and $s' \models \psi$, and (iii) for all s'', if $s \rightsquigarrow s''$ and $s'' \sqsubseteq s'$, then $s' = s''$.

In other words, $(s : \phi)$ is causally relevant to $(s' : \psi)$ just in case: $s \models \phi$, $s' \models \psi$, and s' is a minimal token verifying the relation $s \rightsquigarrow s'$. Thus, mereological minimality comes into the definition of causal relevance twice: first in the definition of the INUS condition (s is an INUS cause of s' just in case s is part of a minimal total cause of s') and, second, in the definition of causal relevance itself.

6. Higher-order Causation without Determinism

6.1. WHY NOT DETERMINISM?

The deterministic model of causation developed in the last section has the advantage of relative simplicity, but there are several reasons for being dissatisfied with such an account of the causal relation.

– Determinism may in fact be false (as on many interpretations of quantum mechanics), and yet it seems clear that the causal relation is realized in our world.
– Thought experiments suggest that a cause need not necessitate its effect. For example, Mackie imagines a machine *L* that indeterministically delivers chocolate bars when a coin is inserted. The machine never delivers bars when no coin is inserted. When a coin is inserted, the machine sometimes does, and sometimes does not, deliver a coin. There is no additional fact that determines which result shall happen on which occasion. When the coin is inserted and a bar is produced, it seems right to say that the coin's insertion was a non-necessitating cause of its effect (Mackie, 1974).
– I have argued in the last section that the causal antecedents of a token are essential to its identity. Consequently, a token-effect necessitates the existence

of its token-causes. If token-causes also necessitated the existence of their effects, each cause-effect pair would be in a relation of mutual necessitation. It would then be very difficult to explain wherein their ontological distinction consisted.

6.2. If not determinism, then what?

The most natural thing to do, if we abandon determinism, is to go probabilistic. We could require that a cause raise the probability of its effect above some fixed threshold (say 90%). We could require that the cause raise the probability of its effect from an infinitesimal to some finite probability. Or, we could simply require that the cause raise the probability of its effect, period.

All such probabilistic relations suffer from the affliction of *nonmonotonicity*. That is, situation s might raise the probability of situation s', but the larger situation $s \sqcup s''$ might lower the probability of s'. Similarly, the type ϕ might raise the probability that the next event will be of type ψ, but the stronger type $\phi \& \mathcal{X}$ might lower that probability. This nonmonotonicity plays havoc with the transitivity of the relation, along with certain desirable connections between causation and statistical generalizations.

The solution, I think, is to talk about *robustly* raising the probability of the effect. A situation s *robustly* raises the probability of s', relative to world w, just in case both s and any extension of s in w raise the probability of s'. Similarly, the pair $(s : \phi)$ robustly raises the probability that ψ situation will follow (relative to w) when both ϕ and the conjunction of ϕ with any other type true of s raise the probability that a ψ situation will follow.

To implement this idea, we would have to introduce a probability measure into our model structures. Instead, I will introduce a qualitative analogue of probability partly because it is somewhat simpler, and partly to establish connections between this theory and existing work on conditional logic (in both the Stalnaker (1981) and Adams (1975) traditions) and the Barwise and Seligman theory of information flow. (Barwise and Seligman, 1997).

In this section, I will introduce a nested set of sets of situations, which I will call a "Lewis system", in order to make clear the connection between this device and Lewis's "system of spheres" semantics for the subjunctive conditionals. In the case of a finite model, each set (or sphere) can be thought of representing the intersection of all those sets whose probability is greater than or equal to $1 - \epsilon$, where ϵ belongs to some fixed order of infinitesimals. In this way, Lewis's system of spheres semantics can be given an interpretation in terms of qualitative probabilities: $\phi \, \mathrm{E} \rightarrow \psi$, represents the condition that the probability of $\phi \& \psi$, is infinitely greater than the probability of $\phi \& \neg \psi$.

The main difference between this use of system-of-spheres semantics and that of David Lewis lies in the absence of the feature of *strong centering*. In Lewis's semantics for the counterfactual conditional, each world is in the center of its own system of spheres. For this reason, the rule of modus ponens is counted as valid for Lewis counterfactuals. When the system-of-spheres semantics is used, as it is in this paper, to represent "fainthearted" conditionals, (Marreau, 1989) the system-of-spheres associated with each situation is not typically centered on that situation. Consequently, modus ponens is not logically valid. Instead, modus ponens counts as a reliable but defeasible rule of inference.

In addition, I am making use of three-valued classification systems, instead of the two-valued interpretations used in traditional modal logic. One consequence of this difference is the failure of the substitution of classically equivalent formulas in modal contexts, including the modal conditional.

6.3. Non-deterministic models

In the indeterministic case, each model shall contain a function $\mathcal{S}^*$ that assigns a Lewis system to each situation-token A *standard, type-invariant model* $\mathcal{M}$ consists of an n-tuple: $< Sit, Typ, \models, \sqsubseteq, \prec, \mathcal{S}^* >$ where:

- *Sit* is a nonempty set, the set of situation-tokens.
- *Typ* is a nonempty set of situation-types, closed under the Boolean operators $\vee$ and $\neg$.
- $\models$ is a binary relation on $Sit \times Typ$.
- $\sqsubseteq$ is a partial, antisymmetric ordering of *Sit*. There is a set $\mathcal{W}$ of maximal situations (worlds). Every situation is extended by some world.
- $\prec$ is a binary relation on *Sit*, and the transitive closure of $\prec$ is a partial ordering of *Sit*.
- $\mathcal{S}^*$ is a function from *Sit* to Lewis systems, where each Lewis system is a nested set of subset of (*Sit*). A classification system corresponds to each member of each Lewis system, when this subset of *Sit* is combined with *Typ* and $\models$.

6.4. Indeterministic causation

In the indeterministic case, we must require that the cause (and not the world or background situation) support the modal conditional linking the cause and the effect:

Definition 6.1 (Dynamic Token Constraint (Indeterministic)). The causal constraint $(s| \sim s')$ holds between tokens s and s' iff s is immediately prior to s' and the optimal s-permitting sphere A in the Lewis system $\mathcal{S}(s)$ for s is such that every world w in A that contains s also contains s'.

If we let Es represent the type corresponding to the actual existence of s(so $s' \models Es$ if and only if $s \sqsubseteq s'$), then we can express this definition by using the fainthearted modal conditional, $\mathrm{E} \rightarrow$:

$$(s| \sim s') \Leftrightarrow (s \prec_0 s') \& s \models (Es \, \mathrm{E} \rightarrow Es')$$

We can now define token causation by making tise of the idea of the *robustness* of a connection in a given situation. Token s supports the causal connection between s' and s'' just in case the causal constraint between s' and s'' is robust in s. In other words, there is in fact a causal connection between s' and s'', and, whenever s' is extended to s_1 in a way compatible with s, there remains a causal constraint between s_1 and s.

Definition 6.2 (Token Causation (Indeterministic)). Token s supports the causal connection ($s' \triangleright s''$) iff both s' and s'' are parts of s, s' is immediately prior to s'', and, for every token s_1 that meets the three conditions – (i) s' is a part of s_1, (ii) s does not exclude s_1 and (iii) s_1 is immediately prior to s'' – the causal constraint ($s_1| \sim s''$) holds.

The definition of causal relevance in the indeterministic case is unchanged from the deterministic case. The definition of causal type/type constraints has to be modified to the following:

Definition 6.3. (Causal Type Constraints (indeterministic)). Token s supports the causal constraint ($\phi| \sim \psi$), between types ϕ and ψ if and only if, in every world w in the optimal ϕ-permitting sphere A in the Lewis system $\mathcal{S}^*(s)$, ϕ tokens (nearly) always support the existence of a causal constraint between themselves and a ψ token.

If the model includes infinitary worlds, then the phrase ‘nearly always’ should be read as meaning: ‘in all but an infinitesimal proportion of the cases’. Once again, the definition of ($\phi| \sim \psi$) can be represented by means of the modal conditional:

$$\exists s'((s' : \phi)\&Es')\mathrm{E} \rightarrow \forall s''\exists s_1(((s'' : \phi)\&Es'') \rightarrow ((s_1 : \psi)\&(s''| \sim s_1)))$$

7. Examples

In this section, I will demonstrate the possibility of higher-order causation. For the sake of simplicity, I will use the deterministic conception of causation throughout. The same examples can be dealt with in a similar fashion, using the indeterministic conception instead.

7.1. Modal and causal facts as causes

Modal facts, such as the fact that ϕ-states are necessarily followed by ψ-states, can themselves act as causes. Consider the tokens s and s', where s is an minimal total cause of s', and s' is a minimal INUS-effect of s (i.e., there is no proper part of s' such that s is an INUS cause of it). Let us assume that s' supports type ψ.

Since s is a minimal total cause of s' and s is a part of itself, we have that $s \rightsquigarrow s'$. Since s' is a minimal INUS effect of s, every type μ such that $s \models \mu$ is causally relevant to the fact ($s' : \psi$).

By the definition of $\triangleright$, s must support the causal constraint $s| \sim s'$. If Hume's Hypothesis applies to this case, then there must be a type ϕ such that s supports both ϕ, and the causal constraint $\phi| \sim \psi$. By the definition of causal relevance (Definition 5.8), we have that the causal-constraint type $\phi| \sim \psi$ supported by s is indeed causally relevant to the explanation of s' and its type ψ. The truth of the causal constraint at s is an indispensable part of the explanation of the actuality of an immediately posterior situation of type ψ.

7.2. First-order teleological causation

Suppose that the fact that wings are causally relevant to flight is part of certain tokens that cause the successful survival and reproduction of a species v of flying bird. The successful survival and reproduction of v is, in turn, causally relevant to the existence of a present-day winged thing, namely, an instance of v. Thus, the existence of an instance of wingedness is explained, in part, by the causal relevance of wingedness to flight. This gives us the initial, Wrightian condition for saying that flight is the function of wingedness as instantiated in this case.

As I mentioned earlier, we can draw a distinction between intrinsic and extrinsic functions.

Suppose that we let ϕ, represent the state of having wings, and ψ the state of flying. Finally, let v represent the entire bird/bird-niche ecological system, including those aspects of the bird's environment that make possible its successful reproduction. The fact that the wings serve the intrinsic purpose of flying in token s can be expressed as:

$$(s' : ((\phi \& v)| \sim \psi)) \rightsquigarrow (s : \phi)$$

The symbol $\rightsquigarrow$ represents the relation of causal relevance, as defined in Section 5.5 above. The state ϕ has the intrinsic purpose of ψ-ing in the token s, relative to background condition v, just in case, the fact that some state-token s' supports a connection between v and ϕ on the one hand, and ψ on the other, is causally relevant to s's being ϕ. In the case of a species v of flying birds, the fact that there is a causal connection between being winged and flying is part of the causal explanation of wingedness in the winged members of v.

In the case of extrinsic purpose, we have instead:

$$(s' : ((\phi \& v)| \sim \psi)) \rightsquigarrow (s : v)$$

In this case, take ϕ to be the presence of suitable seeds in the environment, and take ψ to be the fulfillment of the bird's nutritional needs. In this case, the connection between $v \& \phi$ and ψ causes instances of v, not of ϕ. In other words, the fact that the seeds fulfill the bird's needs explains why there are birds, not why there are seeds. Nonetheless, we can say objectively that, *qua* parts of the bird's ecological niche, the seeds do have the extrinsic purpose of fulfilling their nutritional needs.

Another mode of teleofunctionality is that of representational states, states whose function is to carry information of a certain kind. Elsewhere, I have defined a notion of information, (Koons, 1996) which we can represent by the symbol $\mapsto$. We can say that a particular pattern of retinal stimulation ϕ has the intrinsic function in s (relative to v) of carrying the information that ψ just in case:

$$(s' : [(v\&\phi) \mapsto \psi)]) \rightsquigarrow (s : \phi)$$

The pattern ϕ exists because it carries (in organisms of type v) the information ψ. We might say that when a state occurs that has the function for an organism to carry potential information of a certain kind, then that information has become actual for that organism.

7.3. Higher-order teleology as the basis of mental causation

The efficacy of mental properties depended on the possibility of higher-order teleofunctions. For example, consider the human faculty of inference (whether inductive or deductive). This faculty has the function of interacting with mental states on the basis of their content, a paradigmatically mental or psychological property. Suppose, for example, that mental type ϕ has the function of first recognizing the simultaneous presence of a belief in a conditional and a belief in the antecedent of the conditional, and then producing a new belief (by modus ponens) in the consequent of the conditional. Suppose we have three state tokens, s_1, s_2, and s_3, where s_1 is an instance of the type ϕ, the state whose function is the performance of modus ponens. Suppose that s_2 is a state whose type is that of believing both a particular conditional $(p \rightarrow q)$ and its antecedent, p. Let us call this type of mental state ψ. Finally, let s_3 be a state of believing q (call this type $\mathcal{X}$), immediately posterior to the sum of s_1 and s_2.

We may suppose that the functionality of type ϕ, corresponds to a causal constraint of the form:

$$(\phi\&\psi)| \sim \mathcal{X}$$

Suppose that situation s supports this constraint and contains the sum of s_1 and s_2. We may finally suppose that in the actual world w, s is actually a total cause of s_3. Tokens s_1 and s_2 are both indispensable parts of this cause, and so their mental properties are causally relevant to the outcome. In addition, the fact that mental properties ϕ and ψ, are instantiated can be used in giving a causal explanation of the succeeding state.

It may well be true that tokens s_1, s_2 and s_3 also realize physical states μ_1, μ_2, and μ_3. It may also be the case that the instantiation of μ_1 necessitates the instantiation of ϕ by some super-token, and similarly for μ_2 and ψ, and μ_3 and $\mathcal{X}$. Finally, there may be a covering physical constraint of the form:

$$(\mu_1\&\mu_2)| \sim \mu_3$$

Suppose token s' is a situation containing s_1 and s_2 and supporting this constraint. Then we can suppose that s' is a total cause of s_3. This seems to make the mental properties supported by s_1 and s_2 redundant, otiose. Such a conclusion, however, would however be a mistake. It is true that s' is a total cause of s_3, and that the mental types supported by s_1 and s_2 are irrelevant to the $s' - s_3$ connection. However, it also remains true that s is a total cause of s_3. Token s supports the psychological covering-law but not the physical one. Hence, in the context of s, the **physical properties of s_1 and s_2 are irrelevant, but their psychological properties are not!**

It is true that whenever the organism's behavior can be explained causally in terms of the functional and representational properties of its internal state, the behavior can also be explained solely in terms of the physical and first-order causal properties of that state. There is also some sense in which the first-order explanation is "more fundamental" than the higher-order explanation. However, this fact does not render the explanation in terms of functional terms non-causal, or merely heuristic. Nor does it entail the occurrence of some odd sort of overdetermination. The two explanations do not compete with each other, as two independent physical explanations would do.

Genuine overdetermination requires that one of two conditions be met: (i) the two token-causes of the effect are causally unrelated and mereologically disjoint, or (ii) the relevant types supported by the two causes are logically independent. It would be very problematic to postulate the existence of massive overdetermination of behavior, by both mental and physical causes. However, on my account, behavior is not overdetermined, despite the fact that it caused both on the first-order level (by physical states) and on the higher-order level (by mental and other functional states). In these cases, neither of the two conditions for genuine overdetermination are met.

8. Conclusion

The application of situation theory to the problem of teleological causation has paid dividends in several areas.

– I have been able to reduce teleological relations to a set of causal relations, which in turn can be analyzed in terms of relations of necessity and propensity between situation-tokens.
– I have an explanation for the hyperintensionality of teleological and cognitive contexts, that is, for the fact that these contexts are so sensitive that in many cases, classically equivalent formulas cannot be substituted for one another without altering the truth value of the whole.
– A teleological approach can explain the possibility of false beliefs and other representations, since a mental state can have the function of carrying the information that a certain type is realized, even though no token realizes that type in the actual world.

– I have demonstrated that higher-order causal relations, of the kind required by the Taylor/Wright definition of teleology, are quite possible, constituting a special case of a more general definition of causation, and I have done so without making use of the dubious notion of higher-order objective chance.
– A teleological account of the mind offers a solution to the mystery of mental causation, since mental causation can be understood as involving the operation of higher-order functions, functions taking functionally-characterized states as inputs and outputs. The situation- theoretic account of causation enables us to understand why the existence of multiple total causes at different logical levels does not involve any objectionable form of overdetermination.

There are a number of additional areas where this sort of approach might be extended. For instance, a causal theory of our knowledge of modal facts, including our knowledge of logical and mathematical necessities, is a possibility. The rehabilitation of teleology opens up the possibility of reviving the eudaemonistic tradition of Plato and Aristotle in ethics, identifying the good life as one in which certain kinds of teleofunctions associated with human nature are fulfilled. Finally, this approach enhances the attractiveness of a teleological account of knowledge and warranted belief, along the lines recently proposed by Plantinga (1993).

Notes

[1]This example is due to Anil Gupta, in conversation.
[2]These structures have been independently discovered many times over. Birkhoff (1940) called them "polarities", and Hardegree (1982) called them "contexts". They were also invented by the German mathematician Wille, whose work is discussed in Davey and Priestley (1990).
[3]For a detailed discussion of this connection, see Appendix B of the 1992 paper by Lehmann and Magidor (1992). Adams's book (1975) is the locus classicus of this approach; see also McGee's very insightful paper McGee (1994), and a more recent paper by Morreau (1997) connecting these conditionals with recent work on defeasible reasoning.

References

Adams, E.W. (1975), *The Logic of Conditionals*. D. Reidel, Dordrecht, The Netherlands
Barwise, J. (1989), *The Situation in Logic*, CSLI Lecture Notes. CSLI, Stanford, Calif.
Barwise, J. and Etchemendy, J. (1987), *The Liar: An Essay on Truth and Circularity*. Oxford University Press. Oxford, U.K.
Barwise, J. and Perry, J. (1983), *Situations and Attitudes*, MIT Press, Cambridge, Mass.
Barwise, J. and Seligman, J. (1997), *Information Flow: The Logic of Distributed Systems*. Cambridge, U.K., Cambridge University Press.
Bedau, M. (1992), Where's the good in teleology? *Philosophy and Phenomenological Research*, 52: 781–801.
Bigelow, J. and Pargetter, R. (1987), Functions. *Journal of Philosophy*, 84: 181–196.
Birkhoff, G. (1940), *Lattice Theory*, volume 25 of *Colloquium Publications*. American Mathematical Society, Providence, Rhode Island.
Darwin, C. (1859), *On the Origin of Species by Means of Natural Selection*. Murray, London, U.K.
Davey, B.A. and Priestley, H.A. (1990), *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, U.K.

Davidson, D. (1980), *Essays on Actions and Events*. Clarendon Press, Oxford, U.K.

Dawkins, R. (1987), *The Blind Watchmaker*. W.W. Norton & Co. New York, N.Y.

Dennett, D.C. (1995) *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* Simon & Schuster, New York, N.Y.

Devlin, K. (1991), *Logic and Information*. Cambridge University Press, Cambridge, U.K.

Dretske, F.I. (1981), *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.

Dretske, F.I. (1981), *Explaining Behavior: Reasons in a World of Causes*. MIT Press, Cambridge, MA.

Eells, E. (1991), *Probabilistic Causality*. New York, N.Y., Cambridge University Press.

Gilson, E. (1984), *From Aristotle to Darwin and Back Again: A Journey in Final Causality, Species and Evolution*. University of Notre Dame Press, Notre Dame, IN.

Hardegree, G.M. (1982), An approach to the logic of natural kinds. *Pacific Philosophical Quarterly*, 63: 122–132.

Hitchcock, C. (1996), A theory of second order causation. *Erkenntnis*, 44: 369–77.

Koons, R.C. (1996), Information, representation, and the problem of error, in Jerry Seligman and Dag Westerståhl, (ed.) *Logic, Language and Computation*. Volume 1, pp 333–346. CSLI Publications, Stanford, Calif.

Kripke, S.A. (1972), *Naming and Necessity*. Harvard University Press, Cambridge, MA.

Lehmann, D. and Magidor, M. (1992), What does a conditional data base entail? *Artificial Intelligence*, 55: 1–60.

Leibniz, G.W. (1988), *Metaphysics / Correspondence with Arnauld / Monadology*. Open Court, La Salle, IL.

Lycan, W.G. (1996), *Consciousness and Experience*. MIT Press, Cambridge, MA.

Mackie, J.L. (1965), Causes and conditions. *American Philosophical Quarterly*, 2: 245–264.

Mackie, J.L. (1974), *The Cement of the Universe: A Study of Causation*. Clarendon Press. Oxford, U.K.

McGee, V. (1994), Learning the impossible, in Ellery Eells and Brian Skyrms, (ed.), *Probability and Conditionals: Belief Revision and Rational Decision*, pp 179–199. Cambridge University Press, Cambridge, U.K.

Millikan, R. (1984), *Language, Thought and Other Biological Categories*. MIT Press, Cambridge, MA.

Millikan, R.G. (1989), In defense of proper functions. *Philosophy of Science*, 56: 288–307.

Morreau, M. (1997), Fainthearted conditionals. *Journal of Philosophy*, 94: 187–211.

Neander, K. (1991), Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, 58: 168–184.

Papineau, D. (1993), *Philosophical Naturalism*. Blackwell, Oxford, U.K.

Plantinga, A. *Warrant and Proper Function*. Oxford University Press, Oxford.

Plato, (1986), *The Republic*. Prometheus Books, Buffalo, N.Y.

Sorabji, R. (1964), Biological functions. *Philosophical Quarterly*, 14: 288–302.

Stalnaker, R. (1981), A theory of conditionals, in William L. Harper, Robert Stalnaker, and Glenn Pearce, (ed.), *Ifs: Conditionals, Belief, Decision, Chance and Time*, pp. 41–56. Dordrecht, D. Reidel.

Stampe, D. (1977), Towards a causal theory of linguistic representation. in P. French, T. Uehling, and H. Wettstein, (ed.), *Minnesota Studies in Philosophy*, 2, pages 42–63. University of Minnesota Press, Minneapolis, MI.

Taylor, C. (1964), *The Explanation of Behavior*. Routledge & Kegan Paul, London, U.K.

Woodfield, A. (1976), *Teleology*. Cambridge University Press, Cambridge, U.K.

Wright, L. (1976), *Teleological Explanations*. University of California Press, Berkeley, CA.