Dispositions, Teleology and Reductionism

ALICIA JUARRERO ROQUÉ
Prince George’s Community College

Larry Wright’s *Teleological Explanations*¹ is one of the more recent in a series of discussions on teleology that includes notable works by Braithwaite, Nagel and Charles Taylor.² Unlike Braithwaite and Nagel, Wright argues that teleological ascriptions of behavior qualify as legitimate explanations. Unlike Taylor, he argues that teleological explanations are compatible with mechanistic, neurophysiological accounts of behavior. Wright’s claim, of course, is that his analysis avoids the errors of the other theories. I shall argue that his analysis of teleology fails because it is dispositional and, *a fortiori*, behaviorist. If I am correct, moreover, my objections apply as well to Nagel, Braithwaite and Weiner—indeed, to any other behaviorist analysis of purposive behavior such as Sommerhof’s.³

Section I consists of a brief sketch of Wright’s main thesis. In Section II I argue that Wright’s dispositional analysis of teleological ascriptions is inadequate for several reasons: (1) it cannot take into account behavior that is unique but nevertheless teleological, (2) it cannot accurately identify the precise goal of teleological behavior, and (3) it will often be unable to identify behavior as purposive even though it is. Thus I conclude that Wright’s *analysans* is inadequate.

Wright treats “etiological explanation,” that is, explanation that tells why the behavior came about, as the genus term, with causal and teleological forms constituting the more specific species. His use of “etiology” corresponds to that employed in medical contexts: to provide the etiology of a disease is to identify that which is responsible for the occurrence of the disease. It is Wright’s main contention that since a teleological characterization such as “*A* in order that *B*” explains etiologically “Why *A*?” by saying “something in general about why behavior of this sort takes place” (p. 25), the distinction between teleological and causal explanations “is a distinction among etiologies” (p. 27), not a distinction between etiological explanations and something *sui generis*. 
Classical etiological analyses by Weiner, Braithwaite and Nagel are unsatisfactory, Wright argues, because they cannot accommodate, e.g., behavior such as unsuccessful attempts to flee from an attacker which, although they do not achieve their goal, nevertheless exhibit the teleological characteristic of "trying-behavior." These analyses, moreover, do not recognize that teleological behavior need not be precise. For example, if there are three doors, any one of which would serve as an avenue of escape for an animal fleeing from a predator, it is not necessary that the animal take a given door for his behavior to be described as teleological.

Wright's own analysis of teleological explanation consists of a modification of that presented by Charles Taylor. According to Taylor, behavior is teleological if it occurs because it is required for an end, and explanation of such behavior will be provided by means of "laws in terms of which an event's occurring is held to be dependent on that event's being required for some end." Wright calls this a "requirement etiology" (p. 32); it shows how the "consequences of teleological behavior can function in its own etiology" (pp. 34–35).

Wright argues that Taylor's analysis needs modification because "required" is too stringent a condition. As mentioned earlier, if the animal can flee by either running through a door or jumping over a fence, neither alternative is exactly "required" for the behavior to be an example of goal-directed activity. Wright proposes that we substitute a "consequence etiology" (p. 38) for Taylor's requirement etiology. According to this modified view, behavior is teleological if it occurs because it is the "sort of thing that tends to bring about a certain goal" (p. 38). Thus

"S does B for the sake of G iff
(i) B tends to bring about G
(ii) B occurs (i.e., is brought about by the fact that) it tends to bring about G"

Wright refers to this *analysans* of teleological explanations as (T).

By using mazes, changing the paths within mazes, and so forth, it is possible, according to Wright, to make repeatable and intersubjective tests that will determine whether the behavior satisfies (T). Thus the ascription–explanation that B occurs in order that G can be empirically tested by a procedure which is methodologically akin to that establishing standard causal relationship since it proceeds by eliminating alternative accounts.
Testing whether $B$ exhibits dispositional characteristics enables us to “discover what something is trying to do” (p. 47).

Wright claims that teleological behavior must be analyzed dispositionally but differs from behavior resulting from nonteleological dispositions because the latter are dependent on “straightforward antecedent conditions” (p. 58) whereas the former are less precise and depend on results or consequences. The dependency, however, is “exactly the same: it is the usual sort of causal link” (p. 58).

Wright then argues that dispositional analysis of teleological behavior is not incompatible with mechanism. (T), he notes, is behavioral and, as such, a teleologically characterized $B$ is logically independent of any possible underlying mechanism or internal structure causally responsible for $B$—which is to be expected, he says, if the “underlying mechanism is to explain” $B$ (p. 59). But, it might be objected, if the tendency to bring about $B$ is only a contingent correlate of an underlying mechanism, then the mechanism cannot explain $B$. According to Wright, such an objection “displays a fundamental misunderstanding of both cause and explanation” (p. 61). Although to give a cause is to give an explanation, the explanatoriness of such an answer, he argues, is not necessarily dependent on its identification of a causal factor. The existence of an underlying mechanism does not show that a disposition does not exist; the underlying mechanistic structure would simply provide a contingent guarantee of the dispositional explanation. On Wright’s view the disposition is merely a manifestation of an underlying regularity.

It is sometimes argued that since intentionality is the defining characteristic of action, no additional mechanism can be an adequate ground for explaining action. Consequently, an explanation of human action in terms of an underlying neurophysiological mechanism would not explain action, only movements. Wright, however, believes that one form of mechanism escapes this objection.

[S]uppose we do find a structural state $S$ which manifests itself in a disposition $D$ toward behavior $B$, and that $S$ together with some other parameters $P$ is sufficient for $B$ . . . it is simply not clear whether $S$ and $P$ would still be sufficient for $B$ if things were such that $S$ no longer manifested itself in $D$. [ p. 131]

On this view, then, the structure–behavior relationship would “itself be conditional on some law or rule governing the intention or
purpose” (p. 132); thus it would explain action, not just behavior. Where the agent’s point of view is relevant in explaining human action, the impact of such beliefs, intentions, etc., i.e., what could be characterized as the agent’s “perspective,” can be analyzed dispositionally and thus incorporated into \((T)\). No additional ascription of mental life itself is necessary, Wright claims.

II

Wright’s theses, then, are that \((T)’\)s non-contingent, dispositional account is quite adequate to explain the teleology of human action, that teleological explanations are legitimate forms of explanation, and that maintaining that “wanting to do \(X\)” and “doing \(X\)” (\(ceteris paribus\)) are logically connected is still compatible with a mechanistic, neurophysiological, explanation of human action.

Such a dispositional analysis of teleology places heavy emphasis on its alleged empirical testability. To be “objective enough to function in an experimental test,” conditions must be “repeatable and intersubjective” (p. 45) such that testing for teleological behavior is akin to that which establishes causality. The teleological character of a rat’s searching behavior for food can be tested empirically because one can devise and conduct experiments in which, for example, the pattern of the maze in which the rat is placed is altered, food pellets are replaced by inedible items, and so forth. So too can unsuccessful but nevertheless teleological behavior be detected empirically, as Wright claims, by means of observing whether the behavior exhibits “systematic organization of the movements around a goal-object” (p. 49). But although Wright recognizes that teleological dispositions differ from non-teleological ones, he fails to recognize that all dispositional, and indeed all behavioral analyses of teleology rely on identifications of patterns of behavior that persist through time. This characteristic, however, is neither necessary nor sufficient for teleological behavior.

1. The first condition of \((T)\) requires that for \(S\) to do \(B\) in order that \(G\) it must be the case that \(B\) tends to bring about \(G\). A disposition refers to a tendency to behave a certain way. “Tendency,” an eminently behavioral concept, can be understood only with reference to a pattern of behavior that usually does produce or yield \(G\). Wright claims that the inclusion of this condition accounts for behavior which is “appropriate, given the goal, though unsuccessful” such as the above-mentioned trying–behavior of an animal desperately but unsuccessfully attempting to flee from a predator. Wright fails to recognize, however, that \((T)\) cannot
account for behavior that is teleological and successful in achieving its goal while being inappropriate, given the goal. That is, suppose behavior B never brings about G except just this once; ordinarily B is just the wrong sort of thing to try (in order that G) in these circumstances and S knows it, but this time it succeeds. I must emphasize that G here does not just happen as a fortuitous by-product of B; S does B specifically in order that G. A good example of this might be: S, a student with nothing to lose, flatters his teacher to get a higher grade even though S is well aware that flattering is for the most part the wrong sort of thing to try, and in particular with this teacher. Yet this time it works!

Such a situation is not empirically testable. It is teleological behavior but cannot satisfy the first condition of (T). That is, one cannot demonstrate that B occurred (was performed by S) because it tends to bring about G since, ex hypothesi, B does not tend to bring about G. Dispositional criteria are always inadequate as explanantia of unique situations or uncharacteristic behavior because unique or uncharacteristic behavior does not conform to a pattern and dispositions necessarily imply the existence of a pattern.

2. There is a second type of “uncharacteristic” behavior that (7) cannot account for. Wright’s emphasis on the empirical quality of (7) requires that the subjects under consideration intend a certain goal for at least the time span necessary for tests to be conducted. One can only test for the presence of teleological goal-directedness by examining the behavior which exhibits the same direction over a given length of time. But suppose S never has before and never will again do B—in—order—that—G, but he does this once. S is ordinarily (always has been and always will be again) a peaceful fellow. Just this once, however, in an uncharacteristic fit of anger, he fires his gun fully intending to kill A. If, again, this is a once—in—a—lifetime behavior, no empirical tests could be devised that would determine that he intended to kill A rather than, say, that the gun fired accidentally. There would be no “systematicity” over time around a goal—object that could be tested which would clearly identify the behavior as goal—directed. A variation of this example would be a subject who continuously changes his mind. His behavior, although unintentional and teleological, would exhibit no systematicity around a goal—object.

It might be objected that this second type of case actually meets both (i) and (ii) of (T), and that even though it might be difficult to obtain evidence for (ii), lie detector tests, hypnosis, and other exotic possibilities exist which enable us to test (T) empirically in “unique”
cases where testing via systematicity is barred to us. But such an objection can succeed only by ignoring Wright's repeated claim that the value and advantage of \((T)\) lie in its ability to ignore reference to mental life itself, even for the determination that \(S\) does \(B\) because it tends to bring about \(G\)--(ii).

But do lie detector tests and hypnosis avoid such references? I think not. Lie detector tests and attempts to hypnotize people have been known to fail. In order to rely on such tests, therefore, one must first be able to ascertain each time that the test is administered that it is yielding valid results. And how would one go about doing that without appealing to "mental life" itself? Surely it is obvious that using another (second order) empirical test to determine whether e.g., the lie detector (first order) empirical test is accurate would simply launch us on an infinite regress. Nor could one assume that consistent test results establish the test's validity. Clearly the point being made by my suggested counterexample is that the \textit{validity} of any empirical testing procedure to determine whether \(A\) did \(B\) because of (or in order to) \(C\) must necessarily refer to the mental life of \(A\). It is the assumption that the results of the tests (whether systematicity or hypnosis) accurately report a certain mental state which confers validity on the tests. Not referring to these mental states carries with it the inability to distinguish between accurate and inaccurate tests.

In connection with this example it must also be pointed out that Wright's version of mechanism, which he believes to be compatible with teleological explanation, requires that "\textit{intent be inter alia a disposition}" (p. 134). Even though one might construe this second counterexample in such a way that it seems to meet both conditions of \((T)\) since shooting a gun \textit{does} tend to kill \(A\) and \textit{is} performed because it tends to kill \(A\), the example is formulated such that the intention \textit{responsible for} the behavior is \textit{not} amenable to dispositional analysis in virtue of its once-in-a-lifetime character.

3. A third possibility exists. Suppose that at time \(T_1\) \(S\) does \(B\)--in--order--that--\(G\), says he runs a maze in order to reach food items he genuinely desires at that time. Now let us further suppose that as soon as he completes the first test run he suddenly realizes that he is a subject of a laboratory experiment in which Wright is attempting to test for teleological behavior by means of \((T)\). The teleological end--condition that Wright is testing for, let us suppose, is that \(S\)'s goal is to eat the food \((G)\). Finally, suppose \(S\) is a kindly but dull sort of fellow who likes Wright and wants to please him. As Wright goes about altering the maze and other conditions, \(S\) continues to do \(B\)--"in--order--that--\(G\)," even when he becomes satiated and would
prefer never to see another morsel of food. It seems obvious that $S$ is doing $B$ not just "in-order-to-get-food ($G$)" but rather "in-order-to-get-food-in-order-to-please-Wright ($H$)" but that there would be no way for Wright to discover this latter teleological end-condition empirically.

4. Lastly, imagine that $S$ is now a mischievous character determined to confuse and mislead any experiment of Wright’s. $S$ always knows what Wright is testing for, and methodically does whatever $B$ will not appear to be in-order-that-$G$. Even if Wright becomes suspicious and begins to test whether $S$ is doing $B$-in-order-to-confuse-Wright, $S$ changes his behavior accordingly and makes it appear amenable to testing. Once again, where no consistent pattern of overtly repeated behavior exists, no intersubjective tests can be run and there would be no way to determine that any teleological orientation whatsoever is present or, a fortiori, that the goal is $G$ rather than $H$.

Here it might be objected that these last two cases of experimental deception constitute examples of a well-known problem with an equally well-known solution: carefully disguising the real purpose of the experiment or sometimes even the existence of the experiment makes the in-order-to-please or the in-order-to-mislead reaction impossible. Although a serious practical problem for Wright, it is, the objection might claim, not a theoretical flaw in his dispositional analysis. Once again, however, for this counterobjection and its proposed solution to succeed, one must be able to determine incontrovertibly and without appealing to mental life that the disguise has succeeded on any given trial. But is that possible? Again, one would need to test for the success of the disguise by using empirical techniques only. But, as was the case in the earlier counterexample and for the same reasons, either an infinite regress will be generated or one will find oneself guilty of begging the question. The empirical test which is designed to bypass any reference to the agent’s mental life must, in order to establish its validity, assume that it has accurately reported on that very mental life. And, at the risk of being repetitive, that is exactly what Wright argues need not be done.

Four examples have been presented: in the first case, behavior-type $B$ never (except this once) tends to bring about $G$ so one cannot test for the presence of the first condition of ($T$). In the second, agent $S$ never (except this once) does $B$ in order that $G$. Since his behavior does not exhibit a pattern, no empirical testing as required by ($T$) is possible. In the third case, $S$ deliberately makes his
behavior appear that he is doing B in order that G when in fact he is doing B in order that H. Applying (T) would lead one to conclude that teleology is present but will be in error with regards to precisely what the goal is. Finally, S in the fourth example makes his behavior appear non-teleological even though it is. Applying (T) will not enable one to discover the presence of teleology.

Wright’s disclaimer that “in each case it is possible that we are deceived” (p. 48) does not succeed in salvaging his theory. Part of his criticism of Nagel et al., is that their analyses “cannot accommodate a substantial range of clear and objective goal-directed behavior” (p. 29). The above situations I suggest as counterexamples to Wright’s claims are not so far-fetched that they can be compared to saying, “Well, it is possible that the vase was already broken and just held together by a fortuitous magnetic field that dropped in intensity coincidentally with the jolt” (p. 48). While (T) may give us apparently the best explanation, it sometimes will not give us the correct one—but the traditional teleological explanation will. Thus the analysans does not successfully analyze the analysandum.

Wright realizes that “to demonstrate to a skeptic that a particular B . . . occurs because it is the one that will bring about G . . . requires . . . the elimination of alternative accounts of the phenomenon” (p. 41). Thus he might argue that examples 3 and 4 above are simply off the mark. But how does one eliminate alternative accounts of unique behavior without appealing to “mental life”? Different goals can be pursued by the same sort of behavior and quite different types of behavior can share the same goal. As Nowell Smith points out, granted that someone exhibits dispositional tendencies to act in certain ways,

it is still open to us to ask what is his motive for so acting. His actions are quite consistent with his wanting to gain kudos or his having his eye on the post-war political scene; and they are also consistent with his wanting to help his country. And it is only in this last case that we should call him ‘truly patriotic’ since patriotism consists in doing things for the sake of one’s country. As Aristotle would have said, this ‘for the sake of’ clause is part of the essence of every motive, and it is just this clause that distinguishes a motive explanation from a dispositional explanation.5

If I am correct, in those cases of human behavior in which a mental item (such as a motive identifying a given instance of behavior as
being e.g., in-order-that-G rather than in-order-that-H) is not amenable to a dispositional analysis, alternative explanations cannot be eliminated solely by the behavioral tools of (T). An appeal to these mental items themselves becomes the only way of eliminating alternative accounts. But Wright claims that “mental items are forced on us only as a way of characterizing in an intelligible way the incredible complexity of the (largely teleological) dispositional state of affairs that manifests itself in typical human behavior” (p. 144), and since the impact of mental items can be characterized dispositionally, reference to mental life itself is said to be superfluous.

Wright, however, fails to realize that the impact of two different beliefs, motives, etc., can become manifest in the same dispositional tendency to behave in similar ways. However much one alters the environment etc., the tendency of A and B to behave in similar fashion might remain even though the motivation behind the two behaviors differs. Divergent motives such as gratitude (M1) and enlightened ambition (M2) might result—become manifest—in the same disposition D1 to behave similarly, B1. To use Nowell Smith’s example, the two quite different motives, ambition and patriotism, might become manifest in the dispositional tendency to (a) seek public office, (b) donate large sums of money to charitable organizations, and so forth. It takes no great imagination to construct a scenario in which the dispositions to behave a certain way are indistinguishable even though the two sets of behavior spring from very different intentions. And in this type of case, two different actions will have been performed even though the overt physical movements may be identical. Explaining a specific instance of behavior as that action (and not another) is therefore crucial if the behavior is to be explained at all. And this dispositions alone cannot do, for a disposition is “behavior-oriented,” and “neutral,” so to speak, to the intention from which it issues.

Contrary to Wright’s claim, then, reference to behavioral dispositions alone explains nothing and reference to mental items themselves is necessary in order to distinguish the behaviors. A strictly empirical, behavioral account will not do because dispositions to behave a certain way are not uniquely correlated with mental items.

According to Wright, one appeals to mental items only in order to be able to characterize S’s (otherwise unintelligible) behavior intelligibly. But in situations such as examples 3 and 4 above, how would the experimenter ever know that he needs to do so? How
would he be able to determine that he even needs to appeal to mental items when the subject’s overt behavior seems to provide such a clear case of “systematic arrangement about the goal–object”? Is it not the case that, rather than appealing to mental states in order to explain the incredible complexity of behavioral patterns which one has (antecedently) recognized, to the contrary, one recognizes that there is a “complexity” to be resolved in the behavioral manifestations because one already knows that there is (or might be) a difference in the agent’s motivation, i.e., mental life? One does not infer a difference in mental states in order to characterize differences in the dispositional tendencies themselves because, as the example is set up, there are no empirically detectable differences in the behavior patterns. On the contrary, one can characterize the two examples of behavior as “different” (actions) only because one has independently identified a difference in mental states.

Wright’s particular version of mechanism which he claims is compatible with teleology—a neurophysiological theory which is itself conditional on dispositions—on the one hand seems to require that intent and other mental terms be dispositions (p. 134); on the other hand he specifically states that “none of this entails that, for example, beliefs are dispositions. [As mentioned,] all it requires is that the impact of beliefs (etc.) on action can be characterized in a dispositional way” (p. 149). It is clear from the “etc.” as well as from the footnote on that page that he makes no distinction among mental items such as desire, belief and intent.

Although it should be clear for the reasons given above that I do not agree that teleological behavior can be satisfactorily analyzed dispositionally, let us suppose for present purposes that such an analysis is possible.

Wright’s neurophysiological account is that a certain structural (neurophysiological) state S will “manifest” itself in a disposition D to behavior B; S together with certain parameters P will be sufficient for B. He then adds that

it is simply not clear whether S and P would still be sufficient for B if things were such that S no longer manifested itself in D. If we were to change things so radically that S no longer manifested itself in D, what reason do we have to suppose that we would not thereby change the role of S vis-à-vis sufficient conditions for B? [p. 131]

Although Wright is not very clear here, P seems to identify the
physiological laws in virtue of which the agent is disposed to behave in a certain way.

Problems similar to those above still arise, however. Structural state \(S\) is said to be only contingently related to the disposition \(D\) which is its "manifestation"; thus it is possible that neurophysiological state \(S_1\) can manifest itself in more than one disposition, \(D_1\) and \(D_2\). It is also possible that more than one structural state, \(S_1\) and \(S_2\), can each manifest itself in the same disposition \(D_1\).

Let \(\rightarrow\) signify "manifests itself in," and \(\Rightarrow\) signify "causes."

Consider the first alternative:

\[
(S_1 \rightarrow D_1) \land P \Rightarrow B_1
\]

Suppose \(D_1\) is an "action-disposition," i.e., one connected to consciousness and goal-directedness (intent, belief, etc.); \(B_1\) will, accordingly, be an action. If in the second alternative, \((S_1 \rightarrow D_2) \land P \Rightarrow B_2\), \(D_2\) is not an action-disposition but rather a teleological disposition responsible for changes in physiology designed to maintain homeostasis, behavior pattern \(B_2\) will not be an action. What seems unclear is the explanatoriness of \(S\) with respect to \(B\). It neither explains \(B_1\) as an action (\(D_1\) does), nor does it explain \(B_2\) as not an action (\(D_2\) does). It will be recalled that elsewhere Wright specifically states that the explanatoriness of an answer is not necessarily dependent on the identification of either a cause or a sufficient condition. It seems, then, that \(S_1\) explains nothing at all.

Similarly, in the logically possible case where \(S_1\) and \(S_2\) each manifests itself in \(D_1\):

\[
(S_1 \rightarrow D_1) \land P \Rightarrow B_1\quad \text{and} \quad (S_2 \rightarrow D_1) \land P \Rightarrow B_1
\]

A possible example of this situation is the fact that more than one area of the brain is responsible for the disposition \(D_1\) to do \(B_1\)—speech areas in the brain are various, for instance. Once again, I find it difficult to see what \(S_1\) and \(S_2\) add to the explanatoriness of \(D_1\). Clearly it is not Wright's claim that whereas actions are explained by dispositions (P. 121), dispositions are explained by neurophysiology; his concern is (1) to establish that a structural account can play some role in the explanation of actions, and (2) to show that such a mechanistic explanation is compatible with the explanation provided by the analysans of a teleological characterization.

But this criticism is, perhaps, unfair. He could maintain that there is in fact a one-to-one correlation between each structural state and dispositional tendency; the situations I have presented, then,
simply would not arise. Indeed, such is the claim of neurophysiologist A. Rosenblueth. According to Rosenblueth, each mental event is paired with a neurophysiological correlate. "Each specific mental event," he says, "has as a correlate a specific spatiotemporal pattern of neuronal activity." I do not think that this absolves Wright from criticism, however, for his is not a metaphysical claim. Wright's thesis revolves around the question of explanation. And Rosenblueth won't help him out here. Knowledge of the physiology of nervous centers, he states, "would not allow us to infer the qualitative characteristics of the mental states from those of the nervous activities correlated with them." To support his contention that mental concepts and physical concepts cannot be interchanged, Rosenblueth mentions the standard example of the impossibility of conveying to a blind (from birth) person the qualitative connotation of e.g., "red." The qualitative import that is conveyed by an explanation of an action that makes reference to intent, belief, etc., cannot be conveyed even by a theory which makes neurophysiological explanation conditional on the structural states being manifested in disposition. Thus, even if a mechanistic account were compatible with teleological behavior in the sense that one neuronal pattern were invariably correlated with a specific mental event, the compatibility is independent of the explanatory contribution of each. The qualitative import of explaining that certain behavior was brought about intentionally by a specific mental item is that which identifies the behavior as an action, as a certain action. It is therefore the mental item (or reference thereto) that constitutes the core of the ascription–explanation that is the teleological characterization.

Wright states that the structural account "could contain something in its algorithmic/theoretic details corresponding to 'this is an action–disposition'—but it need not" (my italics, p. 134). Although he does allow that "in the context in which the distinguishing features of these dispositions are interesting" they can be identified, he adds that it is "underlying explainability once again notwithstanding" (my italics, ibid.). But if nothing within a theory distinguishes an action as such then nothing explains it. Likewise, if nothing within a theory explains this action as being in–order–that–G rather than in–order–that–H, then that action has not been explained at all. This defect is particularly crucial in those cases in which the same dispositional tendency to behave a certain way can serve different goals and, therefore, results in either two different actions, or in one action and one behavior pattern that is not an action.
NOTES


7. Ibid., p. 108.