

REVAMPING ACTION THEORY

Gordon Park Stevenson
Trinity College

ABSTRACT: Philosophical interest in intentional action has flourished in recent decades. Typically, action theorists propose necessary and sufficient conditions for a movement's being an action, conditions derived from a conceptual analysis of folk psychological action ascriptions. However, several key doctrinal and methodological features of contemporary action theory are troubling, in particular (i) the insistence that folk psychological kinds like beliefs and desires have neurophysiological correlates, (ii) the assumption that the concept of action is "classical" in structure (making it amenable to definition in terms of necessary and sufficient conditions for its proper application), and (iii) the assumption that deferring to *intuitions* about the application of the concept of action amidst the context of fantastical thought experiments furnishes an effective method for judging the adequacy of proposed analyses. After consideration of these problems it is argued that action theory needs to be reoriented in a more *naturalistic* direction, the methods and aims of which are continuous with those of the empirical sciences. The paper concludes with a sketch (and defense) of the methodological foundations of a naturalistic approach to intentional action.

Key words: intentional action, folk psychology, conceptual analysis, philosophical naturalism, behavior control

For many years intentional action has been a popular subject of philosophical attention, spawning a distinctive branch of philosophy known as "philosophy of action" or "action theory." Action theorists have typically viewed their job as involving conceptual analysis of the language of everyday action ascriptions. This task is undertaken with the assumption that a successful clarification of commonsense action-talk will provide necessary and sufficient conditions for an event's being an action, conditions spelled out in terms of mental kinds lifted from the framework of "folk psychology."

Philosophers and psychologists view folk psychology as the collection of cognitive and linguistic skills that people exploit in their everyday efforts to predict and make sense of one another's behavior. Developmental psychologists have shown that between the ages of three and five, normally acculturated children readily acquire the folk psychological skills of agency ascription via the attribution of desires, beliefs, and personality traits to people and other animals (see White, 1995).

AUTHOR'S NOTE: Thanks to Ruth Millikan for her invaluable help with previous incarnations of the present paper, and also to Armando Machado, George Graham, and an anonymous referee for their constructive criticisms of an earlier draft. Please address all correspondence to Gordon Park Stevenson, Department of Philosophy, Trinity College, Hartford, CT 06106. Email: gpstevenson@sbcglobal.net

Typically, the necessary and sufficient conditions proposed by action theorists involve specifications of the causal relations between the folk psychological kinds deemed requisite for action to occur. These folk psychological kinds include beliefs, desires, intentions, and plans. Furthermore, it is often assumed by action theorists, whether explicitly or implicitly, that the mental kinds invoked in a given definition of action will subsequently be identified or associated with certain kinds of internal (usually neurophysiological) states.

While not all work done under the banner of “philosophy of action” exhibits all of the features just listed, much of it does.¹ Let us henceforth refer to this dominant strain of action theory as GOFPA (Good Old-Fashioned Philosophy of Action). I label it “old-fashioned” because it is basically Aristotelian in character, although it is only since Davidson’s re-articulation of it in the 1960s that it has become something of a dominant tradition among philosophers concerned with agency.

In this paper I critique the agenda of writers in the GOFPA tradition. It is not my aim to summarize and critique exhaustively the work of each and every writer in the tradition, but rather to illustrate that there *is* such a tradition and to highlight some of the tradition’s main doctrinal and methodological features, features which are subsequently challenged later on in the paper. While action theory as it has heretofore been practiced is not a particularly fruitful research program, this should not be taken as cause for despair because GOFPA is *not* the only game in town. An alternative, more *naturalistic* approach to intentional action is possible, and I sketch the methodological foundations of such an approach in the final section of the paper.

Good Old-Fashioned Philosophy of Action

The Aristotelian approach to action aims to explain action from a refined folk psychological perspective, a perspective that takes the ostensible causal flavor of everyday mental kinds like “desire” and “choice” very seriously, treating the referents of such terms as internal states of organisms. Donald Davidson in particular deserves credit for this revival of action–explanation in terms of efficient causation by inner states, states classifiable in terms of kinds derived from conceptual analysis of our everyday explanations of intentional action. Writing during a period in which the Aristotelian view of agency had fallen into some disrepute, Davidson sought to revive the view that mental states like beliefs, intentions, and desires can be seen as legitimate causal antecedents of action.

On Davidson’s early account, what is required for a movement to count as an action is for the movement to be caused by the interaction of a relevant belief/pro-attitude (desire being the principal pro-attitude) complex in the mind of the agent.

¹ e.g. Davidson (1980), Goldman (1976), Brand (1984, 1989), Mele (1992a, 1992b, 1997), Ginet (1990), Odegard (1988), Adams and Mele (1992), Mele & Moser (1994), Wilson (1989), Costa (1987), Audi (1993), Bratman (1987), Davis (1979), and Taylor (1966). There are, of course, many more examples.

REVAMPING ACTION THEORY

This belief/pro-attitude complex constitutes an “intention” or a “primary reason” for action; a movement “done for a reason” or “done intentionally” is an action. As Davidson (1980) characterizes his early view, an action is intentional “if it is caused in the right way by attitudes and beliefs that rationalize it” (p. 87). Davidson argues at some length that mental entities can be token-identified with physical states, and that rationalization—the project of explaining an action in terms of its agent’s underlying mental states—is a species of *causal* explanation.

The type of action theory defended by Davidson, and subsequently by a slew of others, has come to be labeled “causalism,” and it has become the dominant paradigm within action theory. In his introduction to an anthology of essays in action theory, Mele (1997) explicates the main tenets of causalism:

Causal theories of action hold that an event’s being an action depends upon how it was caused. These theories feature as causes such psychological or mental items as beliefs, desires, intentions, and related events (e.g. acquiring an intention to *A now*). . . .Causalism typically is embraced as part of a naturalistic stand on agency according to which mental items that play causal/explanatory roles in action are in some way dependent upon or realized in physical states and events. (pp. 2-3)

Various causalist analyses of action are possible, and a number of writers have proposed alternatives to the Davidsonian causalist framework. However, setting particular theoretical differences aside, most contemporary writers in philosophy of action accept a basic form of causalism according to which, as Horgan and Graham (1990) put it, “an action is an item of behavior involving a certain characteristic kind of causation by [folk psychological] states like belief, desire, and intention” (p. 300).

Adequately specifying the necessary and sufficient conditions for an intentional state to cause a bodily movement *in the “right” way* has proved to be the thorniest analytical problem for causalists. Consideration of the possibility of *deviant causal chains* has led some writers to question whether causation by belief/pro-attitude complexes is indeed a sufficient condition for action. Causal deviancy scenarios involve cases in which the requisite mental states are present in someone, yet the causal path leading from these mental states to the completion of the action in question is unusual enough, due to luck or environmental circumstances, to make one question whether an intentional action has actually been performed. In order to understand the methodology employed by GOFPA theorists, let us focus for a bit upon this causal deviancy problem as it has been confronted by Davidson and another early causal theorist, Alvin Goldman.

Recall Davidson’s causalism, according to which only behaviors caused by *primary reasons* (which are analyzed into belief/pro-attitude complexes) qualify as intentional actions. Goldman independently developed his own form of causalism around the same time as Davidson, and here is an initial definition of action from Goldman (1976):

Act-token A is intentional if and only if (a) there is an act (-type) A' such that the agent S wanted to do (exemplify) A', and (b) either S believed that his doing A would generate his doing A' or S believed that his doing A would be on the same level as his doing A', and (c) this want and this belief caused S's doing A. (pp. 54-55)

Davidson and Goldman each consider some hypothetical cases that threaten to provide counterexamples to their definitions. For example, Goldman imagines a case involving a highly accidental and “intuitively” non-intentional act that nevertheless appears to satisfy conditions (a)-(c) of his analysis. His case: “Suppose a man believes that if he kills his uncle he will inherit a fortune and suppose he desires to inherit a fortune; this belief and desire may so agitate him and cause him to drive in such a way that he accidentally kills his uncle. . .” (p. 55). Davidson (1980) considers a similar counterexample to his analysis: “A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death. Do we want to say that the man killed his victim *intentionally*?” (p. 78). These two cases involve the general causal deviancy problem discussed above, i.e., the relevant pro-attitudes and beliefs, while present in the mind of the agent in question, fail to cause the action *in the right way*. This problem has led some theorists to tighten up their causalist analyses (e.g., Goldman, Brand, Mele) and has led others to soften up their causalism altogether (e.g., Davidson).

Typically, once action–definitions are revised in the light of counterexamples, they are subsequently tested again via the counterexample method until at last the theorist is satisfied that her analysis of intentional action is adequate. This follows the classical methodology of *conceptual analysis*:

- (i) Propose tentative *necessary* and *sufficient* conditions for something being an X (e.g., P is an action *if and only if* it is caused by primary reasons).
- (ii) Devise cases of (what seem *intuitively* to be) Xs that nevertheless seem to violate the conditions proposed in (i) (e.g., Q is caused by primary reasons but does not appear intentional).
- (iii) Revise and tighten up the conditions proposed in (i) in light of the counterexample(s) devised in (ii) (e.g., P is an action *if and only if* it is *proximately* caused by primary reasons).

In response to causal deviancy worries, some writers have sought to tighten up their analyses by stipulating the further condition that actions be *proximately caused* by certain kinds of mental entities, e.g., “occurrent” wants and beliefs (Goldman, 1976) or “immediate here and now” *intendings* or *volitions* (Brand, 1989; Davis, 1979; Sellars, 1973). This follows from the apparent fact that intentions may be causally relevant to an action’s performance without counting as *proximate* causes.

Consideration of the possibility of action in the *absence* of belief or desire, such as appears to be the case with spontaneous or mistakenly performed actions, has led some writers (e.g., Costa, 1987; Davis, 1979; Sellars, 1973) to speak of *volitions* rather than beliefs or desires as being the essential, proximal causal antecedents of action. Costa (1987) has provided a quite involved account of how nested hierarchies of intentions and volitions with varying degrees of content specificity need to be posited in order to define action. Furthermore, differing conceptual analyses of “intentions” have led some writers to *enlarge* the rather frugal catalog of mental entities (i.e., beliefs and desires) with which writers like Davidson, Goldman, and Audi (1993) have made do. Some insist, contra Goldman and Audi, that “intentions” are *not reducible* to belief/desire complexes and propose that in addition to, or in place of, beliefs and pro-attitudes, various sorts of intentions (including “distal,” “guiding” intentions as well as “immediate” intentions) must figure as necessary causes of actions.² Furthermore, some insist that *plans*, as the “representational contents of intentions,” must additionally be invoked in explaining actions (e.g., Mele & Moser, 1994). Lastly, some GOFPA theorists who favor volitions over beliefs and pro-attitudes within their analyses reject the basic tenets of causalism. Prominent non-causalists in philosophy of action have included Ginet (1990), Wilson (1989), and Frankfurt (1978).

Consideration of the above survey of work in GOFPA leads us to a general sketch of the GOFPA explanatory framework. The picture goes something like this: Folk psychology contains an implicit, proto-scientific theory of action which philosophers can make more explicit via the method of conceptual analysis. This folk theory provides a necessary point of departure for further theorizing, and ultimately constrains further scientific theorizing by delineating in a general fashion the sorts of properties psychological or neurophysiological states must exhibit in order to qualify as valid causal precursors to action. Presumably, the mental kinds and relations between mental kinds unearthed by a successful conceptual analysis of folk psychological action-ascriptions will ultimately be identifiable or associable with kinds of inner states and relations among them.

While not every GOFPA theorist would agree with every claim broached in the above sketch, one can discern in it two principal doctrinal/methodological tenets that lie at the core of *all* work in GOFPA (regardless of any particular stance on causalism):

(FP) Folk psychology provides a fruitful conceptual standpoint from which to explain and define the nature of intentional action and in addition serves to constrain subsequent scientific theorizing about action.

(CA) Conceptual analysis provides the theorist with a useful and effective strategy for clarifying and defining folk psychological concepts.

² For more on non-reductive accounts of intention see Bratman (1987) and Mele (1992a, 1997).

There are a number of reasons to be skeptical of (FP) and (CA), hence GOFPA in general, and I enumerate these concerns in the next section of the paper.

Challenging GOFPA

The Problems with (FP): The Internalist Assumption

(FP) has a number of possible avenues of support, namely the “Internalist Assumption” and the “Centrality of Folk Psychology Assumption.” Let’s begin with the former. If the goal of the GOFPA theorist is to mediate the transformation of folk action theory into philosophical (and subsequently scientific) action theory via the clarification and enrichment of folk psychological concepts, then the assumption that folk psychology contains a *proto-scientific theory of action* makes the goal much more achievable; the philosopher’s goal becomes that of transforming what is *already* a rudimentary scientific theory into a more sophisticated and precise theory in line with various scientific models.

There has been a long-standing dispute in the philosophical literature concerning the status of folk psychology as a body of proto-scientific theory. (This debate should *not* be confused with the debate within psychology between “Theory-Theorists” and Simulation theorists concerning what occurs inside a person’s mind during intentional state-ascription.) A number of writers in the past have challenged the causal—and hence, proto-scientific—status of everyday explanation in terms of propositional attitudes (e.g., Dennett, 1987; Melden, 1961; Ryle, 1949; Wittgenstein, 1958); in turn, many others have attacked these anti-causalist arguments (e.g., Churchland, 1989; Davidson, 1980; Goldman, 1976).

My goal in this section is to challenge the assumption that folk psychology constitutes a *certain kind* of proto-scientific theory, one whose posits (e.g., beliefs, desires) will eventually be identifiable or associable with kinds of *internal* states, events, or relations of interest to neurophysiologists and/or cognitive psychologists. (Note that I have not built the *type/token* distinction into the statement of the Internalist Assumption here because GOFPA theorists typically do not exploit that distinction themselves.) This assumption, which I will refer to as the “Internalist Assumption,” is an assumption that most—if not all—GOFPA theorists entertain and, if true, would provide a powerful avenue of support for (FP).

“Internalism” is a label that has been applied by various writers to the Sellarsian view that folk psychology, like neurophysiology and cognitive psychology, posits unobservable inner states in an effort to explain and predict observable behavior. Some writers (e.g., the Churchlands) have stalwartly championed the Internalist vision of folk psychology while simultaneously arguing that folk psychology is a radically *false* body of proto-neurophysiological theory.

(We might label this position “False Internalism.”) In recent times, Internalism in general has come under quite a bit of attack.³

The Internalist Assumption is found throughout the GOFPA literature. Consider some passages from Goldman (1976):

Perhaps the identification of neuro-physiological states that correlate with wanting and believing might help us achieve techniques for measuring wants and beliefs, but certainly at present we do not have this information. (p. 73)

[N]europhysiological information can help explain how it is that wants and beliefs cause action. (p. 167)

Why do wants tend to cause acts? . . . This is a perfectly legitimate question. And perhaps the answer is that wants are correlated with closed loop systems in the cortex, the firing of which causes efferent neuron firings which in turn cause muscle movements and limb movements. (pp. 168-169)

In Mele (1992a) we find the assumption that *intentions*, individuated by their content, can in general have “neural realizations,” and furthermore that *distinct* intentions—individuated by their distinct contents—have *distinct* neural realizations. Of a certain woman he asks us to imagine that

She had the intention, N, of her opening the window, “that by it she would” let in some fresh air; and she had the intention, O, of her opening the window, that by it she would gain a better view of the street. Suppose that a neuroscientist, without altering the neural realization of N itself, renders that realization incapable of having any effect on S’s bodily movements. . . while allowing the neural realization of O to figure normally in the production of movements involved in S’s opening of the window. . . (p. 212)

Elsewhere, Mele (1997) suggests that certain “neurophysiological effects of the acquisition of [an] intention. . . *realize* [a man’s] trying to raise his arm. . .” (p. 16) and, in general, many GOFPA theorists appear optimistic about the likelihood of—as Odegard (1988) aptly puts it—a “neurophysiological extension” of GOFPA doctrine.

Unfortunately, there is an important reason to be skeptical of the Internalist Assumption, namely that *content individuation* within folk psychology does not appear *fine-grained* enough to facilitate the mapping of (tokens of) folk psychological kinds onto (tokens of) neurophysiological or cognitive psychological kinds. Stich has used such a criticism against a view much like the Internalist Assumption by highlighting cases where one folk psychologically-individuated state would map onto more than one neurophysiological state. For example, Stich (1996) writes:

³ The view called “Internalism” has been challenged by a number of authors. McDonough (1991), Haldane (1988), Wilkes (1991), Baker (1995) and Hornsby (1997) present particularly sustained criticisms of it (Baker calls it “The Standard View”).

There are a variety of psychological dimensions on which people can differ enormously and still be classified by folk psychology as having beliefs that share the same content . . . Some people have sharp vision, others see poorly, and still others are blind. Yet there are circumstances in which folk psychology would attribute the belief that the traffic light has just turned green to all three sorts of people. . . . So for many propositions, it looks like the class of mental states that folk psychology will count as having content will be very heterogeneous indeed. The neurological states subserving these beliefs will differ drastically, both physically and functionally. (p. 26)

Writers have also stressed the *converse* of this point, i.e., that there will be cases in which more than one folk psychologically-individuated state maps onto a single (global) neurophysiological state. For example, Baker (1995) asserts that in order for a view like the Internalist Assumption to gather evidential support,

. . . a scientific theory would have to find a relevant neurological difference between a brain state that constitutes a belief that p and a brain state that constitutes a belief that q for any two distinct beliefs. Such a theory must be able to distinguish between a brain state that constitutes . . . say, a belief that a soldier's following orders is a slightly mitigating factor in assessing misconduct and a [a brain state that constitutes a] belief that a soldier's following orders is a substantially mitigating factor in assessing misconduct. (p. 16)

It is naïve to expect there to be a distinct brain state behind each and every distinct attribution of a folk psychological state. Even a casual consideration of our everyday practices of belief and desire attribution reveals the surprising extent to which folk psychological states exhibit “substrate neutrality,” being reasonably attributable to both normal and mentally challenged humans, monkeys, dogs, etc. When we are dealing with folk psychology, we need to be honest about the way that we conceptualize things *qua folk*, and *qua folk* we are all more than willing to ascribe thoughts, beliefs, desires, and intentions to all sorts of humans and animals that are not capable of sharing identical kinds of neurophysiological states. This is not to say that folk psychological kind attributions have *no* connection whatsoever with the presence of certain kinds of internal states in organisms, but rather that we cannot safely assume that the tokening of a certain folk psychological kind uniquely picks out a token neurophysiological state.

Apropos of this is the “Missing Innards” thought experiment (first devised by Stich, 1983) involving a scenario in which we encounter a being that appears to have beliefs and desires but which has either no innards or extremely simple ones. It then seems to follow that given knowledge of these missing innards, we would likely retract our original intentional ascriptions to the creature, and this reveals that in fact “what’s in there *does* matter somehow” (Clark, 1993, p. 216). Now, if belief ascriptions are *not* posits of inner states, as I have suggested thus far, then awareness that a believer lacks any inner states should *not* count as evidence against our ascribing beliefs to her.

The Non-Internalist reply to this seeming problem involves the concession that while being a creature capable of belief *in general* may contingently require a

certain degree of internal complexity, having a particular belief B versus having a particular belief C need not hinge upon differences in internal structure. In other words, while being deemed a creature to whom we may suitably attribute beliefs may hinge upon perceived similarities between the creature's internal structure and ours, folk psychological kinds themselves need not be posits of discrete kinds of internal states. The latter simply does not follow. But then again, the intuition fueling this "problem" may be at fault; given sufficiently intelligent behavior by a creature with missing innards, we might very well go on ascribing beliefs to the creature. Given such behavior, it would presumably not even occur to us *qua folk* to get a closer look at the innards. (Only the scientists among us would be at all interested in such an investigation, and only the Internalists among us would actually look for *beliefs* within the innards.)

The defender of the Internalist Assumption might respond to the "wide-grained-content" objection by insisting that discrete folk psychologically-defined mental states might ultimately be associated with *sets* of internal states rather than discrete internal states. This is a view welcomed with open arms by some functionalists in philosophy of mind. For example, Horgan and Woodward (1985) concede that folk psychological "events might well be identical with arbitrarily complex, highly gerrymandered [brain] events which themselves are not naturally isolable relative to [cognitive science]" (pp. 411-412).

This response, though, is problematic since the neurophysiological states subserving a given belief might differ *functionally* as well as physically (recall Stich's visual belief example). This would entail that the complex sets of internal states onto which a folk psychologically-characterized state would be mapped might be seen as highly baroque and uninteresting from *both* a neurophysiological *and* a psychological point of view, and not simply from a neurophysiological point of view as some functionalists (e.g., Fodor, 1981) have suggested. The Internalist Assumption would remain dubious because it claims that folk psychological kinds will be identifiable or associable with kinds of internal states, events, or relations *of interest to scientists in general*. If folk psychological kinds ultimately do not prove to be of interest to scientists in general, then the validity of (FP), and the GOFPA project at large, is severely threatened.

Some writers have gone further than the present content-width objection and insist that the ascription of folk psychological states is a *strongly indeterminate* affair. In other words, it is claimed that given a single frame of reference, it is not the case that a single proposition is *uniquely* believed by a person. As Dennett (1991) puts it,

there could be two different systems of belief attribution to an individual which differed *substantially* in what they attributed—even in yielding substantially different predictions of the individual's future behavior—and yet where no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not. (p. 49)

Dennett's position is indebted to Quine's "repudiation theory," according to which facts about the manner in which we learn the application of mental

concepts—most importantly, the fact that we learn mental concepts via publically observable behavioral signs alone—entail the *nonexistence* of mental entities. Quine convincingly championed the radical indeterminacy of folk psychological state ascription, and accordingly underscored the futility of any attempt to reduce folk psychological kinds to kinds of interest to neuroscientists (e.g., see Quine, 1960, 1985; Stemmer, 2001). Such Quinean indeterminacy considerations provide yet another line of attack against the Internalist Assumption.

If the Internalist Assumption were false, then one might wonder what alternative vision of folk psychology makes better sense. Several writers have defended what we might label an “Externalist” view of folk psychology. According to this Externalist view, nicely encapsulated by Clark (1989),

. . . folk psychology is *designed* to be insensitive to any differences in states of the head that do not issue in differences of quite coarse-grained behavior. It papers over differences between individuals and even over differences between species. It does so because its purpose is to provide a general framework in which gross patterns in the behavior of many other well-adapted beings may be identified and exploited. The failure of folk psychology to fix on, say, neurophysiologically well-defined states of human beings is thus a virtue, not a vice. (p. 48)

It is certainly possible that what we qua folk psychologists are “tracking” are observable patterns of behavior rather than unobservable internal states. The real targets of folk psychological ascriptions need not be (and probably aren’t) psychological states in any “narrow” sense.

Some writers have concluded that folk psychology is not characterizable as a body of scientific theory of *any kind* due to the fact that it “fails to postulate easily individuated entities” (Christensen & Turner, 1993, p. xix) and involves “undefined or underdefined theoretical terms and other features which would be inexcusably lax in any *scientific* offering” (Preston, 1989, p. 291). However, these statements go too far because it remains a live possibility that folk psychology comprises an “Externalist” body of proto-scientific theory targeting wider behavioral patterns or historical/causal relationships rather than internal states alone.

The Problems with (FP): The Centrality of Folk Psychology Assumption

A second avenue of support for (FP) is the notion that action *must*, to a large extent, be explained from the perspective of folk psychology because action is originally defined at that level. Call this the “Centrality of Folk Psychology Assumption.” The Centrality of Folk Psychology Assumption does not necessarily involve the idea that action can *only* be explained from the folk psychological level. This stronger assumption can be found in the work of certain writers, however, and we might refer to it as the “Only Game In Town Assumption.” According to the Only Game In Town Assumption, when it comes to explaining action, folk psychology provides the *only* framework in town; once we leave folk

psychology we change the subject and no longer analyze *action*. Writers avowing the Only Game In Town Assumption (e.g., Baker, 1995; Hornsby, 1997) are not in the GOFPA mainstream because they reject the Internalist Assumption, yet their position needs to be questioned along with the more moderate Centrality of Folk Psychology Assumption.

While the concept of intentional action is ultimately derived from the practice of folk psychology, it does not follow that any adequate explanation of intentional action must take place largely at the folk psychological level or that the scientific explanation of action must be grounded in folk psychological explanation to any theoretically significant degree. An analogy may help to illustrate this point. Consider our everyday concept of water. This concept originated at the folk level, we might presume, to serve entirely practical purposes, yet it would seem odd to insist that since this concept originated at the everyday explanatory level, any adequate explanatory analysis of water must be framed in terms of concepts available solely at that level. On the contrary, it would seem that an adequate explanation of water is best served by looking *beyond* the everyday framework that has introduced the concept and by seeking theoretical concepts situated at the ontological levels of chemistry and physics.

In general, there is no reason to believe that a concept can best, or *only*, be explained from the framework in which the term originates. The Only Game In Town Assumption is particularly misguided in this respect. Claiming that a non-folk psychological or “sub-personal” account is not in fact concerned with *action* (but, rather with brain states and/or bodily movements) simply because a sub-personal account doesn’t conceptualize actions in the same way as does folk psychology is analogous to saying that a chemist does not study *water* since his concept of that stuff is quite different from the concept of that stuff we entertain when viewing it from an everyday perspective. The conclusion does not follow in either case.

It is unclear what sort of semantic theory is invoked by espousers of the Only Game In Town Assumption. While the everyday concept of water may not be *identical* to the modern chemical concept of water as H₂O, there is an important sense in which the chemical concept amounts to an *enrichment* of the everyday concept rather than the realization of an entirely new concept. Perhaps only a realist, externalist semantics like that argued for by Putnam (1975) can allow us to make sense of this important point. According to such a semantics, the commonsensical concepts lurking behind our everyday tokenings of the terms “person” and “action” can be seen as picking out or latching onto spatiotemporal parts of the world that a scientific approach can subsequently investigate in more detail. In this way, our everyday concepts of “mind” and “action” might function as *placeholders* for further scientific analysis (more on this in the paper’s final section).

As suggested above, an adequate explanation of the extension–set of a given concept often requires looking *beyond* the framework that has introduced the concept and seeking novel theoretical concepts and models situated at different levels of analysis than that occupied by the framework which introduced the

concept. This point applies especially in the case of folk psychological concepts such as intentional action because it is likely that these were not self-consciously introduced as theoretical terms targeting the ontological level of brain states. However, it *would* be strange to suggest that a scientific account of action might proceed without *any* consideration of the folk psychological concept. The Centrality of Folk Psychology Assumption hinges upon the intuitively reasonable idea that although the framework introducing a concept C may not have exclusive “ownership rights” over C, it nevertheless places *some* constraints upon how C is to be explained or how the range of C is to be expanded to include novel extensions. Typically, GOFPA theorists view such constraints as *theoretical* constraints, i.e., particular constraints on how a scientific theory is to explain the causation of actions. For example, Brand (1984) claims that folk psychology constrains the scientific psychological analysis of action by requiring that all actions be caused by proximate intentions having a distinctive “conative” component; scientific psychologists are then left with the task of “transforming” the folk notion of proximate, “conative” intention into a more empirically acceptable theoretical construct.

According to another perspective, folk psychology may place *some* constraints on any naturalization of a folk psychological concept, but only to a *very limited* extent. These constraints will be limited for the simple reason that folk psychology likely does not constitute a proto-scientific body of neurophysiological or cognitive psychological theory, as argued above. If folk psychology can be seen as placing any constraints on the scientific analysis of action, these constraints will be semantic in nature: folk psychology may constrain scientific theory by stipulating, in effect, that a certain class of extensions comprises the prototypical set of actions and that any sorts of extensions not in this prototypical set must be shown to be similar in fundamental ways to the members of this set before it can be allowed that the novel cases are in fact cases of *actions*. (By “prototypical set of actions” is meant simply those cases that are non-controversially agreed to constitute actions by a majority of successful wielders of the concept.) Note that whatever we determine to be the “fundamental” similarities between prototypical actions and newly encountered cases cannot itself be presumed to be given to us *by folk psychology* since it is possible that folk psychology itself contains no unified theory concerning how to apply the concept of action to new or previously encountered instances of behavior.

Even if folk psychology initially places some semantic constraints on scientific theorizing, it is likely that the scientific analysis of action will result in the development of completely novel concepts unrecognizable from the folk psychological perspective. This has already taken place in the case of *learning* and *memory*—two concepts which (like action) were originally developed within the framework of folk psychology and whose meanings have subsequently been transformed by contemporary clinical neuroscience, psychology, and artificial intelligence. We might agree with the idea behind the Centrality of Folk Psychology Assumption that folk psychology places *some* constraints on scientific theorizing about action without agreeing that these constraints need be

theoretically robust. Accordingly, it remains questionable to require, as does the Centrality of Folk Psychology Assumption, that the theorist keep one foot firmly in the realm of folk psychology as she endeavors to understand and explain intentional action in her chosen manner.

The Problems with (CA): The Classical Definition Assumption

It should be clear from my survey of work in GOFPA that theorists in the tradition depend heavily upon the methodology of conceptual analysis. Accordingly, it is natural to attribute to them the assumption that conceptual analysis provides us with a useful and effective procedure for the clarification and definition of folk psychological concepts. I have labeled this assumption (CA).

Why would anyone believe (CA)? What other assumptions might warrant it? There are two possible avenues of support for (CA) that GOFPA theorists must at least tacitly accept if (CA) is to appear rationally motivated. The first avenue of support I will call the “Classical Definition Assumption” and the second I will call the “Fruitfulness of Intuitions Assumption.” Let’s begin with the first of these.

For conceptual analysis to be a useful and effective methodology for the explication of folk psychological concepts, folk psychological concepts must be amenable to explication via the methodology of conceptual analysis, and for the latter to be the case it must also be true that folk psychological concepts are susceptible to classical definitions in terms of necessary and sufficient conditions for their proper application. (This follows from the fact that there *is* a characteristic methodology associated with the process of conceptual analysis, namely the three-step methodology enumerated in the first section of the paper.) This latter claim I will refer to as the “Classical Definition Assumption,” and it targets the *structure* of folk psychological concepts.

Some writers, most explicitly Brand, have conceded that entirely successful conceptual analyses of folk psychological notions will probably not be forthcoming due to the vagueness and non-systematicity of our everyday intuitions. As Brand (1984) himself puts it, his analysis does “not purport to capture the totality of pre-analytic intuitions,” since “those intuitions, more than likely, do not form a consistent, unified whole” (p. 141). So, on the surface, someone like Brand might not appear to accept the Classical Definition Assumption. However, this does not stop Brand—or any other GOFPA theorist—from altering his analytic strategy one bit, and any theorist who knowingly and explicitly does conceptual analysis in the manner of GOFPA theorists must be committed to something like the Classical Definition Assumption. Otherwise we would have to accuse such a person of irrationality—why analyze concepts if you ultimately believe they are not amenable to the sort of analysis you undertake?

Brand’s response might be to the following: Where everyday intuitions provide no clear, consistent guidance concerning the application of concepts to hypothetical cases, enrichment of our folk psychological framework according to the demands of science will be necessary, the goal not being an exhaustive conceptual analysis *per se* but rather a *partially successful* conceptual analysis that

reveals the basic outlines of a theory of action. However, this response prompts further difficult questions regarding the aims of GOFPA theorists. For example, what constitutes a “partially successful” conceptual analysis? And might the improbability of a *fully* successful conceptual analysis suggest some deeper problem with the methodology of conceptual analysis itself?

In asserting that folk psychological concepts like intention and action are susceptible to definitions in terms of necessary and sufficient conditions, the Classical Definition Assumption depends upon what has come to be referred to as the “classical” view of concepts. According to the classical view, concepts have the structure of definitions determining the proper application of the terms expressing them, and definition of a concept can be represented in terms of singly necessary and jointly sufficient conditions (for more on this matter see Bishop, 1992). Hence, any challenge to the classical view of concepts will also constitute a challenge to the Classical Definition Assumption.

There have indeed been challenges to the classical view of concepts within psychology and philosophy over the past few decades. One of the more well known challenges came out of the work of Eleanor Rosch, whose research several decades ago showed that there are “typicality effects” associated with everyday subjects’ categorization of subordinate kinds under superordinate kinds that cannot be easily explained by the classical view. Rosch’s prototype theory, according to which concepts involve a weighted list of prototypical category feature representations, has come under attack in recent times, however, and there are now a wide range of alternate theories of concept structure on the market (for a helpful summary of problems with the prototype theory, and for an in-depth discussion of the range of concept structure theories, see Laurence & Margolis, 1999).

Clearly, questions concerning the structure of concepts are ultimately for scientific psychologists to answer, and all that needs to be emphasized here is that there *are* alternative views of concept structure out there, hence it is not defensible to *assume* (without argument) that our ordinary folk psychological concepts do in fact have the structure of classical definitions. If any GOFPA theorists do hold that the everyday concept of action is non-classically structured they have not been explicit either in their avowal of this point or in their disavowal of the classical methodology of conceptual analysis that runs rampant in the literature. Furthermore, no theorist has satisfactorily explained how the proposal of necessary and sufficient conditions for action from the armchair will help further our understanding if the concept of action indeed has a prototype or exemplar-based structure. If the concept of action has a prototype or exemplar-based structure, then empirical methods will need to be exploited in order to understand the nature of the concept; for example, consider the recently popular interdisciplinary, cross-cultural research into the nature of our “folk biological” concepts discussed at length in Medin and Atran (1999).

There is also a *prima facie* reason to be wary of the Classical Definition Assumption, namely that the classical view of concepts does not do justice to the “fuzzy” nature of the everyday concept of action and the variety of “borderline” cases and gray areas associated with the concept. There are quintessential cases of

action and non-action that most of us would agree upon, and then there are the many instances of behavior that we simply are not sure whether or not to characterize as actions. As the history of GOFPA reveals, intelligent, able wielders of folk psychological concepts can disagree sharply over how to apply the concept of action in various cases. If concepts can adequately be represented by necessary and sufficient conditions then how does the fuzziness emerge?⁴ In general, it might be that the classical view of conceptual structure is too rigid to account for the fluidity of folk psychological categorization, and this would explain the failure of GOFPA theorists to achieve consensus on the definition of action.

The Problems with (CA): The Fruitfulness of Intuitions Assumption

Writers in the GOFPA tradition rely heavily upon *intuitions* concerning whether or not folk concepts can appropriately be applied to hypothetical scenarios. Hence, we can reasonably attribute to GOFPA theorists the assumption that deferral to intuitions amidst the process of applying folk psychological concepts to hypothetical cases provides an effective method for judging the adequacy of proposed definitions. Call this assumption the “Fruitfulness of Intuitions Assumption.” The Fruitfulness of Intuitions Assumption provides another potential avenue of support for (CA) and is distinct from the Classical Definition Assumption because unlike the latter assumption, which targets the *structure* of folk psychological concepts, the Fruitfulness of Intuitions Assumption targets the efficacy of GOFPA methodology in *getting at* that structure.

Certain writers in the tradition can be seen as implicitly challenging the Fruitfulness of Intuitions Assumption while simultaneously avowing it in practice. Once more note Brand’s (1984) concession that our folk psychological “intuitions, more than likely, do not form a consistent, unified whole” (p. 141). Yet, in what sense will deferral to intuitions furnish an *effective* procedure for judging the validity of conceptual analyses if it is acknowledged that intuitions about concept application inevitably *contradict* one another?

While this problem arises in the context of rather mundane cases of concept application, it emerges with even greater vigor in the context of the blatantly fantastic hypothetical scenarios upon which the analyses of so many GOFPA practitioners depend. How reliable will our intuitions prove in cases involving omniscient demons, crafty neuroscientists, and the like? How are intuitions expected to provide authoritative answers in cases so far removed from the everyday milieu that has nurtured their development? How can we hope to achieve consensus regarding what our intuitions tell us in such science fiction cases when we can’t even achieve consensus regarding what our intuitions tell us in more mundane cases of folk psychological concept ascription?

⁴ Mele and Moser (1994) try to account for this fuzziness by building vagueness into their proposed necessary and sufficient conditions. Such attempts to accommodate fuzziness by explicitly building in vagueness to definitions are ad hoc and ignore the live possibility that such fuzziness might constitute a clue that our folk concepts do not have the structure that theorists in the GOFPA tradition have for so long assumed.

The fact that our intuitions vary so widely concerning the application of our folk psychological concepts should serve not as an incentive to “tighten up” our pre-analytic intuitions, but rather as a warning sign that our folk psychological concepts might be equivocal, “fuzzy,” and/or non-classically structured. In general, there is currently no reason to accept the Fruitfulness of Intuitions Assumption. However, this does not entail that intuitions serve *no* useful role in philosophical theorizing. Indeed, as Laurence and Margolis (2003) have recently emphasized (in accordance with the view espoused in the present paper), intuitions “can be helpful in orienting an investigation into [a] kind’s nature, especially in the early stages when there is little else to work with” (p. 278). Intuitions concerning the nature of action provide essential (yet fallible) launching points for further investigation, even if they do not constitute fruitful arbiters for testing rival theories within the sphere of conceptual analysis.

Toward a Naturalized Action Theory

Action theory is currently in need of some major revamping. The foundational methodological and ontological assumptions of GOFPA—the preoccupation with classical conceptual analysis, the application of intuitions to fantastical thought experiments, and the over-reliance upon folk psychological kinds—have impeded significant progress. There is a need to reorient action theory in a more *naturalistic* direction, a direction whose methods and aims are continuous with those of the empirical sciences.

One of the goals of “naturalizing” action will be to make it clearer how scientific approaches and strategies from various fields might contribute to our understanding of action. This involves a spirit of inclusiveness. There is a particular need for such inclusiveness because the science of the mind is still in its early stages of development. It could be argued that, at the present time, there is not even enough consensus on foundational questions like “What kinds of things are to count as having minds?” and “Are minds identical to brains?” to allow us to say that there *is* a unitary science of the mind. However, there *are* some promising empirical approaches to such phenomena as movement control and goal-directedness, and a consideration of such empirical approaches might help to enrich a more general philosophical model of intentional action.

There are a number of varieties of philosophical naturalism, hence a number of distinct ways of naturalizing action. For example, Dan Wegner’s (2002) form of naturalism rejects the notion that there is genuine *agency* in the world. Accordingly, he disavows the search for some set of real, scientifically respectable properties of movements that distinguish actions from non-actions. Instead, in a move very reminiscent of Hume, Wegner naturalizes action by analyzing the causal antecedents of our really existent, yet ultimately unfounded and illusory *feelings* of volition and conscious will. On his account, it only *feels* as if there are genuine (i.e., consciously willed) actions in the world, while in reality each person’s behavior is caused largely by unconscious sets of physiological and

psychological processes, processes that we are largely unaware of from a phenomenological point of view.

The main problem with Wegner's otherwise useful theoretical perspective is that it far too narrowly equates "genuine agency" with *consciously willed* action, hence it overstates the case against agency in general. Since Darwin it has become clearer that human cognitive and behavioral capacities, hence human *agency*, are seldom if ever *sui generis*; as Frankfurt (1978) succinctly put it, the "concept of human action is no more than a special case of another concept whose range is much wider" (p. 52). This more "widely ranging" phenomenon—agency *in general*—rather than some putative, empirically problematic faculty of conscious will, should comprise the proper target of a naturalized action theory. Agency *in general*—the capacity to generate purposive behavior—indisputably exists in the world and is exhibited to varying degrees in the behavior of a wide range of biological (and perhaps even artificial) creatures. Accordingly, it does not require consciousness or awareness of agency as such to exist in the mind of the agent. (I defend this point about consciousness in more detail in Stevenson, 2001.)

Given these points, the pressing question becomes "what sorts of scientifically interesting properties distinguish paradigm cases of non-actions from paradigm cases of actions?" The guiding assumption is that there *are* noteworthy differences between actions and non-actions in general, although these will likely be matters of degree, so we might presume that all behaviors lie on a degree-of-agency continuum. (Note how the everyday distinction between action and non-action plays a role here in setting the stage for naturalistic inquiry; more on this shortly.)

Wegner's approach involves one species of naturalism, and it has been found wanting insofar as it defines action too narrowly. Let's consider another species of naturalism implicit in the work of Dretske (1988) and Millikan (1984, 1993) that we might call "Constructive Naturalism." According to Constructive Naturalism, the process of naturalizing action will involve the deployment of *novel* empirically-grounded concepts which taken together may be viewed as *counterpart*-concepts to the folk psychological concept of intentional action. Such counterpart-concepts serve neither as *replacements* for the everyday concept of intentional action nor as *clarifications* of it. The everyday concept of intentional action simply gets the ball rolling by picking out the *tip* of an explanatory iceberg. The rest of the explanatory iceberg will include kinds of behaviors not necessarily recognizable as actions through the lens of folk psychology but recognizable as instances of a scientific counterpart-concept to action in virtue of general properties—referred to as "Underlying Properties"—that the behaviors mutually exhibit and share with behaviors deemed to be paradigmatic human actions from a folk psychological perspective.

Underlying Properties run the gamut from internal structural properties to relational properties, etiological/historical properties, "emergent" physical properties, and abstract mathematical properties. They are kinds that are of interest to empirical scientists and should not be viewed as "essences." It is better to view Underlying Properties as comprising general modes available to the theorist of characterizing and individuating Underlying Kinds independently from the

methods of categorization and individuation we exploit as folk psychologists for the everyday purpose of behavior prediction and explanation.

The intersection of these Underlying Properties delineates the range of an “Underlying Kind,” i.e., the extension-set of a scientifically interesting counterpart-concept to the everyday concept of action. Our folk psychological concept of action might target an Underlying Kind that is more widely ranging than we, *qua folk*, might think. In fact, there are Darwinian reasons to suspect that all instances of cognition and behavior come in varying degrees of complexity, hence they can be located on a continuum relative to one another. The upshot of this Darwinian working hypothesis is that the typical “paradigm” cases of intentional actions by human agents will not be seen as *sui generis*, but instead as comprising an evolutionarily advanced subspecies of a more wide-ranging biological phenomenon. The goal of naturalizing intentional action on the Constructive Naturalist’s approach will be an improved understanding of the broader Underlying Kind (call it “Action*”) of which quintessential human action is simply a subspecies. The ultimate goal will *not* be to analyze the folk psychological concept of action itself with the aid of the novel concepts deployed in the naturalization; *that* would just amount to a fancy form of conceptual analysis.

It may seem as though Constructive Naturalist methodology is simply a form of conceptual analysis in disguise. For example, Tye (1994) has characterized the approaches to naturalization taken by Dretske, Fodor, and others as involving “a priori conceptual reflection” in the service of discovering the *essences* of intentional states. Tye is rightfully pessimistic that “we will succeed in discovering a detailed essence for the property of having some intentional content by *a priori* reflection alone” (p. 138). However, is he right in characterizing the Constructive Naturalist as an *a priori* conceptual analyst?

There is some confusion here concerning the distinction between the “*a priori*” and the empirical, a confusion which may lead one to draw an overly sharp distinction between (i) scientific and (ii) naturalistic philosophical methodology. On Tye’s account, it seems, any sort of theorizing not *directly rooted* in rigorous empirical research counts as *a priori*. However, much of scientific theorizing itself is not *directly* rooted in such research. Consider once more the “naturalization” of our everyday concept of water. One can not tell from armchair conceptual analysis that a central underlying property of water is that of having the particular chemical structure H₂O; this can only be ascertained via rigorous empirical study. Yet water is not reducible to H₂O *tout court*. A whole body of theory, e.g., that visible reality is composed of smaller bits, some of which are invisible to the naked eye, and that these bits can “bond” together in various kinds of ways, etc., must be in place before it even *makes sense* to naturalize water as H₂O, and such general theories are simply the result of *confronting broad patterns in a wide-ranging base of non-rigorous empirical observations*.

The methodology underlying the formation of such general hypotheses in the early stages of scientific inquiry is continuous with the methodology underlying the Constructive Naturalist’s proposal of Underlying Properties. The Constructive

REVAMPING ACTION THEORY

Naturalist starts from a very broad base of non-rigorous empirical data, namely the set of prototypical extensions of a given folk psychological concept. Just because the Constructive Naturalist is also a folk psychologist in everyday life does not rule out the *empirical* nature of her observation base. Since she is not after an analysis of how she, *qua folk psychologist*, happens to determine the application of folk psychological categories, she is not a conceptual analyst. What she is doing is continuous with what all theorists do in the early stages of inquiry: she is trying to find out what sorts of perspectives will be relevant in trying to understand more deeply those aspects of reality we (dimly) refer to by means of folk concepts, and she does this by positing Underlying Properties that she thinks will help us to target patterns and continuities among a broad base of empirical observations. It is simply not “given,” especially in the case of our folk psychological concepts, *which kinds* of scientific perspectives will be pertinent to the task of naturalization. Determining what levels of scientific analysis will be relevant to the study of action (or any other mentalistic concept originating in folk psychology) is perhaps the most challenging part of the naturalist’s task.

What might be the Underlying Properties of Action*? An involved consideration of this question takes us beyond the scope of the present work, but I will conclude by considering the hypothesis that the phenomenon of *behavior control* is a central Underlying Property of Action*. According to this view, actions are movements over which agents are capable of exerting a high degree of control, and non-actions are movements that agents are not capable of controlling to a high degree. Without a doubt, control is something that comes in different degrees, and it appears to be difficult to draw a sharp line between movements of which a creature is in control and those of which it is not. Furthermore, there appears to be no *unitary self* from a neurophysiological point of view. Given these facts, we need to jettison the notion of there being any hard and fast distinction between self-controlled movements and non-self-controlled movements. A more fruitful approach will involve the development of principled methods for placing behaviors on a cross-species control *continuum*. This task, I submit, will require the deployment of models of control from psychology, neurophysiology, and situated robotics as well as an ethologically grounded methodology of behavioral analysis.

Let’s begin by considering some influential models of behavior control from psychology and neurophysiology. For the most part, empirical work on the nature of behavior control has tended to identify control with an agent’s ability to direct its behavior successfully toward some functionally defined end-point or goal in the face of threatening environmental disturbances. Witness the influential work of cyberneticists like Wiener (1948) and Ashby (1956) in developing *closed-loop* or *negative feedback* models of behavior control. Powers (1989) helpfully summarizes the structural elements of the closed-loop model of control:

A sensor reports the state of the controlled variable as a correspondingly variable signal, inside the control system. The signal is compared against a reference signal carried inside the system, and the discrepancy is represented by

still another signal, the error signal. The error signal is amplified to produce a physical output, which in turn acts on the same controlled variable. (p. 27)

Although it has proven to be very fruitful for explaining behavior control in a wide array of cases, the closed-loop model has some noteworthy empirical limitations, e.g., its inability to cover cases in which subjects exhibit control over rapid ballistic movements having durations shorter than possible feedback times (see Lashley, 1917) and clinical de-afferentation cases in which subjects continue to produce goal-directed movements in the absence of *any* afferent information (see Berman & Berman, 1973; Bizzi et. al., 1978). Such scenarios reveal that for certain classes of seemingly controlled, goal-directed movements, the closed-loop model is inadequate.

Concern over the adequacy of the closed-loop model has historically fueled the development of open-loop, *feedforward* analyses of behavior control. These typically involve the notion of a *motor program*, nicely characterized by Abernethy and Sparrow (1992) as “a pre-structured set of centrally stored specific efferent commands which, when executed. . .[allow]. . . a desired movement pattern to be produced without reliance upon ongoing sensory information” (p. 21). The feedforward model easily accommodates species of control that could not effectively be modeled in terms of negative feedback, and some writers have capitalized on this by developing *hybrid* models of control combining features of both open- and closed-loop models within a single framework. The most comprehensive example of such a hybrid approach is Gallistel’s (1980), which remains a very useful account.

The open- and closed-loop models of behavior control address crucial aspects of the phenomenon of control, hence agency. It seems undeniable that an important feature of intentional action is its goal-directedness, and open- and closed-loop models together present helpful ways of understanding how goal-directed behavior control might be physically realized in agents. However, there is an important aspect of behavior control in principle distinct from goal-directedness that most empirical theorists have failed to consider. This neglected aspect of control involves the ability agents have to act differently than how they have in fact acted on a particular occasion, the ability to “snap out” of habitual motor routines. A creature capable of some intricate goal-directed behavior might nevertheless appear, after close empirical scrutiny, to be unable to break out of this behavior once it has been initiated, and this reveals that the creature’s degree of control over this behavior is somewhat limited. There is a sense in which it “could not have done otherwise” than how it in fact acted. (In such a case we might say that the behavior is in control of the agent rather than the other way around!) This potentially important “could have done otherwise” dimension of the phenomenon of behavior control has been widely discussed by free-will theorists and modal logicians but has not yet received much consideration from a *naturalistic* perspective (the notable exception being Dennett, 1984, 2003). In the remaining pages I will sketch the beginnings of a naturalistic analysis of this “could have done otherwise” species of behavior control.

REVAMPING ACTION THEORY

The capacity for behavior control can be viewed from two ontological levels: (i) the ethological level of whole agents in interaction with their environments, and (ii) the physiological or “design stance” level of internal control mechanisms in interaction with one another. We must begin by approaching control in the “could have done otherwise” sense from the level of whole organisms in interaction with their environments. Enlisting the aid of some examples and concepts from the field of ethology, I have elsewhere (Stevenson, 2001) developed the theoretical notion of a “choice point schema” that allows us (in principle) to begin to locate behaviors on a cross-species control continuum. Choice point schemata are blueprints of the temporal loci amidst behavioral sequences at which it is determined via ethological methods of observation that agents have the capacity to interrupt, alter, or prematurely terminate their behaviors. There are four key temporal contexts for control in this interruptability/alterability sense:

- (a) initiation of the movement sequence
- (b) termination of the movement sequence and transition to a new sequence
- (c) moment-to-moment transitions between distinct movements within the sequence
- (d) overall duration and tempo of the movement sequence (temporal aspects of the movement sequence that can be altered)

In a nutshell, the more “choice points” (i.e., temporal loci in a movement sequence allowing for novel movement transitions, alterations, initiations, and/or premature interruptions or cessations of the behavior) a behavior exhibits amidst these four key temporal contexts for control, the greater the extent to which the behavior is under the agent’s control. Note that on this approach one cannot simply decide from the armchair the degree of control exerted by an agent over its behavior; ethologically inspired assessment strategies are required. Let’s illustrate with the example of hunting behavior in the bee-hunting digger wasp (*Philanthus triangulum*) as discussed by Tinbergen (1952):

A hunting female of this species flies from flower to flower in search of a bee. In this phase she is entirely indifferent to the scent of bees: a concealed bee, or even a score of them put out of sight into an open tube so that the odour escaping from it is clearly discernible even for the human nose, fails to attract her attention. Any visual stimulus supplied by a moving object of approximately the right size, whether it be a small fly, a large bumble bee, or a honey bee, releases the first reaction. The wasp at once turns her head to the quarry and takes a position at about 10-15 cm. to leeward of it, hovering in the air like a syrphid fly. Experiments with dummies show that from now on the wasp is very susceptible to bee-scent. Dummies that do not have bee-odour are at once abandoned, but those dummies that have the right scent release the second reaction of the chain. The second reaction is a flash-like leap to seize the bee.

STEVENSON

The third reaction, the actual delivery of the sting, cannot be released by these dummies and is dependent on new stimuli, probably of a tactile nature. (p. 47)

Once we have broken the behavior down into its four apparent phases—search, approach, seizure, and sting—we might sketch out a representation of the *choice point schema* of the wasp’s hunting behavior, with asterisked numbers representing the key choice points in the behavior’s progression:

Search → → → → Approach → → → → Seizure → → → → Sting
1* 2* 3* 4* 5*

In this case, there are five key choice points to consider:

- 1* Initiation of the Search Phase is not directly triggered by a simple sign stimulus but is rather a complex motivational result of both internal drive states (e.g., hunger) and perception of appropriate environmental cues (e.g., sunlight levels), so there is some flexibility regarding the contexts in which the phase can be initiated.
- 2* The Search Phase can be interrupted if a predator stimulus is presented to the wasp, which triggers a transition to an Escape behavior.
- 3* Early on in the Approach Phase, a predator stimulus can, again, trigger a premature cessation of the hunting behavior, but after this point the experimental presentation of the predator stimulus triggers no interruptions.
- 4* The wasp’s encountering a non-bee-like odor in the prey triggers an interruption of the Approach phase.
- 5* If the wasp does not detect a sufficiently bee-like tactile stimulus in the prey, the Seizure phase will be interrupted. Once the Sting phase begins, there are no further temporal contexts in which the wasp can interrupt the ensuing behavior.

It is important to note that choice point schemata involve ethological-level, relatively substrate-neutral maps of the potential for behavior control (in the “could have done otherwise” sense) exhibited by a particular agent amidst a particular kind of behavior. The focus is on the structure of a given behavior in itself, not the wetware that makes the behavior possible. While an ethological level analysis of behavior control is needed so that an agent’s movements can initially be located on a control continuum, the causal antecedents of behavior control reside largely *inside* of the agent, in the agent’s system of internal control mechanisms. It follows that any adequate account of behavior control will also have to consider the phenomenon from the *physiological* or *design stance* level of internal control mechanisms in mutual interaction.

At the internal mechanism level, I have elsewhere proposed a theory stressing the important relationship between “modulatory interconnectivity” and an agent’s

capacity for self-control. Modulatory interconnectivity is a matter of the extent to which an agent's internal "behavior modules"—internal mechanisms that have either been selected for or designed (in the case of robotic agents) because of their capacity to contribute to the production of certain kinds of behavior—are interconnected via links of *inhibition* or *suppression*. In a nutshell, the more informational connections between an agent's behavior control modules that facilitate the modulation of the activity of one module by another, the more the agent will have the capacity to interrupt its behavior short of its typical end-point, to alter its behavior's overall duration and tempo, and/or to initiate or refrain from initiating the behavior in the first place. Since these three capacities are constitutive of control in the "could have done otherwise" sense, we might say that modulatory interconnectivity is a prerequisite for behavior control, hence intentional action.

Although full discussion of this matter is well beyond the scope of the present paper, it is worth noting that the modulatory interconnectivity thesis coheres nicely with some current neurophysiological frameworks that explain human motor control in terms of the extensive inhibitory pathways linking the cerebellum, motor cortex, basal ganglia, and spinal cord (e.g., see Kolb & Whishaw, 1996 for some representative neurophysiological accounts of motor control in line with the present discussion of modulatory interconnectivity). Since we tend to view *human* agents as the species of agents having the highest capacity for behavior control (hence action), a natural way of explaining this developed capacity for control would be in terms of some feature that the human central nervous system exhibits to a greater extent than that of any other biological entity, and if the current thesis is on target, this general feature is modulatory interconnectivity.

To conclude, the primary goal of the final section of this paper has been to elucidate and defend the methodological foundations of a naturalistic approach to action. A genuinely naturalistic account of action will eschew folk psychological conceptual analysis and the application of commonsense intuitions to fantastical thought experiments and will instead aim to develop models of counterpart action concepts such as behavior control that are rooted in and continuous with work in the empirical sciences. At this early stage in the development of cognitive science, the potential fruitfulness of a naturalistic approach to action remains an open question, yet it is certainly time to begin theorizing about action in a fresh manner and to see where it leads us.

References

- Abernethy, B., & Sparrow, W. A. (1992). The rise and fall of dominant paradigms in motor behaviour research. In J. Summers (Ed.), *Approaches to the study of motor control and learning* (pp. 3-46). Amsterdam: North-Holland.
- Adams, F., & Mele, A. (1992). The intention/volition debate. *Canadian Journal of Philosophy*, 22, 323-338.
- Ashby, R. (1956). *Introduction to cybernetics*. New York: Wiley.
- Audi, R. (1993). *Action, intention, and reason*. Ithaca, NY: Cornell University Press.
- Baker, L. R. (1995). *Explaining attitudes*. New York: Cambridge University Press.

- Berman, A. J., & Berman, D. (1973). Fetal deafferentation: the ontogenesis of movement in the absence of peripheral feedback. *Experimental Neurology*, 38, 170-176.
- Bishop, M. (1992). The possibility of conceptual clarity in philosophy. *American Philosophical Quarterly*, 29, 267-277.
- Bizzzi, E., Dev, P. Morasso, P., & Polit, A. (1978). Effect of load disturbances during centrally initiated movements. *Journal of Neurophysiology*, 41, 542-556.
- Brand, M. (1984). *Intending and acting*. Cambridge, MA: MIT Press.
- Brand, M. (1989). Proximate causation of action. In J. Tomberlin (Ed.), *Philosophical perspectives 3* (pp. 137-164). Atascadero, CA: Ridgeview.
- Bratman, M. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Christensen, S., & Turner, D. (1993). Introduction. In Christensen and Turner (Eds.), *Folk psychology and the philosophy of mind* (pp. i-xxv). Hillsdale, NJ: Erlbaum.
- Churchland, P. M. (1989). Folk psychology and the explanation of human behavior. In Christensen and Turner (Eds.), *Folk psychology and the philosophy of mind* (pp. 247-262). Hillsdale, NJ: Erlbaum.
- Clark, A. (1989). *Microcognition*. Cambridge, MA: MIT Press.
- Clark, A. (1993). *Associative engines: connectionism, concepts and representational change*. Cambridge, MA: MIT Press.
- Costa, M. (1987). Causal theories of action. *Canadian Journal of Philosophy*, 17, 831-854.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Davis, L. (1979). *Theory of action*. Englewood Cliffs, NJ: Prentice-Hall.
- Dennett, D. (1984). *Elbow room*. Cambridge, MA: MIT Press.
- Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Dennett, D. (1991). Real patterns. *Journal of Philosophy*, 89, 27-51.
- Dennett, D. (2003). *Freedom evolves*. New York: Viking.
- Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press.
- Frankfurt, H. (1978). The problem of action. In A. Mele (Ed.), *The philosophy of action* (pp. 42-52). New York: Oxford University Press.
- Gallistel, C. (1980). *The organization of action*. Hillsdale, NJ: Erlbaum.
- Ginet, C. (1990). *On action*. Cambridge: Cambridge University Press.
- Goldman, A. (1976). *A theory of human action*. Englewood Cliffs, NJ: Prentice-Hall.
- Haldane, S. (1988). Understanding folk. In Christensen and Turner (Eds.), *Folk psychology and the philosophy of mind* (pp. 263-287). Hillsdale, NJ: Erlbaum.
- Horgan, T., & Graham, G. (1990). In defense of southern fundamentalism. In Christensen and Turner (Eds.), *Folk psychology and the philosophy of mind* (pp. 288-311). Hillsdale, NJ: Erlbaum.
- Horgan, T., & Woodward, J. (1985). Folk psychology is here to stay. In Christensen and Turner (Eds.), *Folk psychology and the philosophy of mind* (pp. 144-166). Hillsdale, NJ: Erlbaum.
- Hornsby, J. (1997). *Simple mindedness: In defense of naive naturalism in the philosophy of mind*. Cambridge, MA: Harvard University Press.
- Kolb, B., & Whishaw, I. (1996). *Fundamentals of human neuropsychology*. New York: Freeman.
- Lashley, K. (1917). The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology*, 43, 169-194.
- Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis and S. Laurence (Eds.), *Concepts* (pp. 3-81). Cambridge, MA: MIT Press.

REVAMPING ACTION THEORY

- Laurence, S., & Margolis, E. (2003). Concepts and conceptual analysis. *Philosophy and Phenomenological Research*, 67, 253-282.
- McDonough, R. (1991). A culturalist account of folk psychology. In J. D. Greenwood (Ed.), *The future of folk psychology: intentionality and cognitive science* (pp. 263-288). Cambridge: Cambridge University Press.
- Medin, D., & Atran, S. (1999). *Folkbiology*. Cambridge, MA: MIT Press.
- Melden, A. (1961). *Free action*. London: Routledge and Kegan Paul.
- Mele, A. (1992a). Recent work on intentional action. *American Philosophical Quarterly*, 29, 199-217.
- Mele, A. (1992b). *Springs of action*. New York: Oxford University Press.
- Mele, A. (1997). Introduction. In A. Mele (Ed.), *The philosophy of action* (pp. 1-26). New York: Oxford University Press.
- Mele, A., & Moser, P. (1994). Intentional action. In A. Mele (Ed.), *The philosophy of action* (pp. 223-255). New York: Oxford University Press.
- Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
- Odegard, D. (1988). Volition and action. *American Philosophical Quarterly*, 25, 141-151.
- Powers, W. (1989). Volition: A semi-scientific essay. In W. Hershberger (Ed.), *Volitional action* (pp. 21-37). Amsterdam: North-Holland.
- Preston, J. (1989). Folk psychology as theory or practice? *Inquiry*, 32, 277-303.
- Putnam, H. (1975). The meaning of "meaning." In K. Gunderson (Ed.), *Language, mind, and knowledge* (pp. 131-193). Minneapolis: University of Minnesota Press.
- Quine, W. V. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine, W. V. (1985). States of mind. *Journal of Philosophy*, 83, 5-8.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Sellars, W. (1973). Action and events. *Nous*, 7, 179-202.
- Stemmer, N. (2001). The mind-body problem and Quine's repudiation theory. *Behavior and Philosophy*, 29, 187-202.
- Stevenson, G. (2001). *The naturalistic foundations of intentional action*. Ph.D. dissertation (unpublished), University of Connecticut.
- Stich, S. (1983). *From folk psychology to cognitive science: the case against belief*. Cambridge, MA: MIT Press.
- Stich, S. (1996). *Deconstructing the mind*. New York: Oxford University Press.
- Taylor, R. (1966). *Action and purpose*. Englewood Cliffs, NJ: Prentice-Hall.
- Tinbergen, N. (1952). *The study of instinct*. London: Oxford University Press.
- Tye, M. (1994). Naturalism and the problem of intentionality. In P. French et. al. (Eds.), *Philosophical naturalism* (pp. 122-142). Notre Dame: University of Notre Dame Press.
- Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- White, P. (1995). *The understanding of causation and the prediction of action: from infancy to adulthood*. Hillsdale, NJ: Erlbaum.
- Wiener, N. (1948). *Cybernetics, or control and communication in the animal and in the machine*. New York: Wiley.
- Wilkes, K. (1991). The relationship between scientific psychology and common sense psychology. *Synthese*, 89, 15-39.
- Wilson, G. (1989). *The intentionality of human action*. Stanford: Stanford University Press.
- Wittgenstein, L. (1958). *Philosophical investigations*. New York: Macmillan.

