Author's Note

This article is based on a talk presented at the 1st Amazonian Symposium on Behavior Research and Theory (Universidade Federal do Pará, Belém, Brasil, November 2013). Correspondence concerning this article should be sent to Marcus Bentes de Carvalho Neto, Universidade Federal do Pará, Núcleo de Teoria e Pesquisa do Comportamento, Rua Augusto Corrêa, 01 Guamá, Belém, PA 66075-110, BRASIL. E-mail: marcusbentesufpa@gmail.com
ABSTRACT: Behavior analysis exemplifies a highly peculiar type of explanation in which behavior is said to arise from past interactions with the environment rather than from internal mental states. Radical behaviorism has been advanced as a philosophy of science that could legitimize this explanatory specificity of behavior analysis. As described by its proponents, radical behaviorism is a philosophy of science that stresses prediction and control, that takes empirical regularities to be the building blocks of behavior science, and that proceeds without hypothesizing unobservable states of affairs. In this article I argue firstly that it is a mistake to ground behaviorism (of any kind) in a philosophy of science, secondly that the philosophy of science known as radical behaviorism is deficient, and thirdly that one can reject this philosophy and make use of a particular type of hypothetical constructs without ceasing to be radical or behavioristic. The philosophy of science I propose as an alternative to Skinnerian inductivism embraces causation as a metaphysical primitive and redirects behavior analysis toward the study of invisible relations of influence between environment and behavior.

Key words: behavior analysis, radical behaviorism, philosophy of science, causation, induction, theory
the boundaries of the behavior-analytic community (Hineline, 1980) are blurred and open to a variety of scientific perspectives. The first and main outlet for basic behavior-analytic research, the *Journal of the Experimental Analysis of Behavior*, has certainly published a number of articles with a cognitivist orientation (e.g., Eichenbaum & Fortin, 2005; Gibbon & Church, 1992; Straub, Seidenberg, Bever, & Terrace, 1979; Williams, 1984). Within the subset of the behavior-analytic community that identifies itself most closely with operant behaviorism, however, the explanation of behavior assumes highly distinctive features. Whereas psychologists traditionally explain performance in terms of internal mental states, operant behaviorists most commonly explain it in terms of interactions between the animal and its surroundings. Another characteristic of operant behaviorism is its heavy appeal to histories of past interactions, as distinguished from stimuli that are close in time to current responding. As Hineline (1990) put it:

An environment-based account, such as that introduced by Skinner, gives a more salient role to immediately eliciting or occasioning stimuli; however, the primary environments of environment-based theory are past environments, for the roles of the present stimuli are seen as dependent upon the organism’s prior history. Even an insensitivity of behavior to immediately attendant stimuli is attributed to past history. (p. 306)

My question in the present article is whether these explanatory practices—the behaviorist’s environmentalism and its reliance on temporally extended histories—have a deeper philosophical justification, and if so, of which kind. Why do behavior analysts spend so much time studying environmental interactions over time instead of appealing to internal mental, or even neural, events? Should the justification, if any, be found in a particular philosophy of science? Or should it be found elsewhere, for example in a set of specific assumptions about the nature of psychological events (e.g., Ryle, 1949)?

**Radical Behaviorism as a Philosophy of Science**

Radical behaviorism is a rationale for behavior analysis that combines general prescriptions about doing science with detailed assumptions about behavior and private psychological phenomena (e.g., Moore, 2008). These assumptions cover topics such as the importance of operant reinforcement, its similarities to natural selection, and its role in verbal behavior. Prescriptions about scientific research loom large in radical behaviorism, however, and many behavior analysts seem to think of their behaviorism as a philosophy of science *sui generis*—a philosophy of science applied to a particular discipline, to be sure, but a philosophy of science nonetheless. Skinner (1969), for example, writes that radical behaviorism is “a philosophy of science concerned with the subject matter and methods of psychology” (p. 221). In the series foreword to Baum’s (1994) *Understanding Behaviorism*, we find that “behavior analysis is deeply rooted in the philosophy of science known as radical behaviorism” (p. xiii), and the book’s second chapter is titled, “Behaviorism as Philosophy of Science” (p. 17). Similarly, Chiesa’s (1994)
introduction to Radical behaviorism: A Distinct Philosophy of Science" (p. 6). In what follows I will draw mainly on Chiesa's (1992, 1994) work because it is the most thorough, among behavior analysts, in articulating radical behaviorism as a philosophy of science.

A first characteristic of radical behaviorism that Chiesa (1992, 1994) emphasizes is its rejection of causation as a metaphysical tie underlying observed regularities. Instead, radical behaviorism adopts a regularity theory of causation of Humean and Machian heritage (Hume, 1777/1966; Mach, 1893/1960) in which metaphysical ties are replaced by “functional relations” (Chiesa, 1992, p. 1290); that is, purely mathematical relations or correlations among events. This aspect of radical behaviorism is explicit in the published works at the origin of behavior analysis, for example in Skinner’s (1931/1972) treatment of the concept of reflex:

We may now take that more humble view of explanation and causation which seems to have been first suggested by Mach and is now a common characteristic of scientific thought, wherein, in a word, explanation is reduced to description and the notion of function substituted for that of causation. (pp. 448-449)

and in some of his comments in The Behavior of Organisms:

In general, the notion of a reflex is to be emptied of any connotation of the active ‘push’ of the stimulus. The terms refer here to correlated entities, and to nothing more. All implications of dynamism and all metaphorical and figurative definitions should be avoided as far as possible. (Skinner, 1938, p. 21)

Skinner’s remarks on causation in Science and Human Behavior (1953) are along similar lines. He writes that the terms of cause and effect are “no longer widely used in science” and that the concept of causal connection has been replaced by that of “functional relation,” meaning merely that “different events tend to occur together in a certain order” (p. 23). In philosophy, the claim that modern science should embrace mathematical regularities instead of metaphysical causation has been advanced most famously by Russell (1912-1913), but it can also be found in positivist books of the late nineteenth century such as Karl Pearson’s (1900) The Grammar of Science (p. 130). Chiesa (1994) endorses this Humean philosophical tradition when she writes that “scientific equations refer to events as a function of other events rather than in terms of A exerting a force on B” (p. 107).

A second prominent characteristic of radical behaviorism as a philosophy of science is its strong inductivist bent and concomitant rejection of hypothetical constructs of any kind. According to Chiesa (1994), the inductive approach of radical behaviorism “attempts to derive general theoretical principles from data” (p. 53) instead of deriving predictions from hypothetical mechanisms. The first stage in the radical behaviorist strategy is to describe functional relations between the environment (in the form of antecedent or consequent stimuli) and behavior. Once these first-order functional relations are identified, the next stage is to use their formal properties as dependent variables which can enter into functional relations with other features of the environment or behavior (Nevin, 1984). This
process can be applied recursively, generating layers of variables of increasing abstraction. This explanatory hierarchy of functional relations is what radical behaviorists consider a theory of behavior (Chiesa, 1992). The observability of first-order functional relations extends to the upper layers, however, so that regardless of their level of abstraction, in radical behaviorism theoretical terms “always describe observed regularities” (Chiesa, 1992, p. 1295).

Other prominent features of radical behaviorism include a technological stance toward science (L. D. Smith, 1992) and a distrust of realism in any of its forms (e.g., Baum, 1994). Perhaps because Skinner’s (1938) research program has been the most successful so far among those that can be considered behavioristic, this inductivist conception of science has become associated with the rejection of mentalism. Conversely, proponents of cognitivism in psychology have been prone to criticize the radical behaviorist philosophy of science and to defend the study of cognition through scientific strategies that differ sharply from the inductivism of radical behaviorists (e.g., Baars, 1986).

In McMullin’s (1984) contrast between what he calls nomothetic and retrodictive explanations, for example, radical behaviorists stand with the former and cognitivists typically side with the latter. McMullin defines a nomothetic explanation as one “that appeals to an empirical regularity, or combination of such regularities” (p. 209). Within the nomothetic ideal of science, the way to explain an empirical regularity is to “move to a higher-order empirical regularity” (p. 201); note the similarity between the nomothetic strategy and radical behaviorism as described by Chiesa (1994). By contrast, the retroductive strategy consists in postulating “entities, properties, processes, relations, themselves unobserved, that are held to be causally responsible for the empirical regularities to be explained” (McMullin, 1984, p. 210). The understanding that retroductive explanation affords “is not just a matter of correct prediction and technical control; it is an opening up of a hitherto hidden world of processes and structures both macroscopic and microscopic” (McMullin, 1978, p. 145).

In psychology, retroductive explanation is consistent with the postulation of internal mental states as causes of overt performance. It is almost a platitude that cognitivist explanations of behavior rely on hypothetical constructs (e.g., Baars, 1986), and indeed, the philosophers who defend scientific explanation through underlying mechanisms explicitly recommend this strategy for cognitive science (e.g., Bechtel & Abrahamsen, 2005, 2010). Even in behavior analysis, the theorists who endorse a more cognitive approach to animal performance defend retroductive explanation and the use of hypothetical constructs over strict Skinnerian inductivism (e.g., Killeen, 1984; Williams, 1986).

What we see, then, is a close correlation between a type of explanation of behavior that is distinguished by its specific ontology (namely, the nature of its explanatory components: past environments instead of current internal states) and a type of explanation that is distinguished by its scientific research strategy (empiricist, inductivist, Humean with respect to causation, and opposed to the use of hypothetico-deductivism. See Tulving (1986).
of hypothetical constructs). Agreement with one type of explanation is taken to imply agreement with respect to the other type.

**From Alliance To Divorce**

A broadly Humean theory of causation, combined with a pragmatic stance on prediction and control, can undoubtedly be taken as a justification for focusing on the environment as an explanans. Perhaps it is for this reason that radical behaviorism, conceived as a philosophy of science (e.g., Moore, 2003), is popular among behavior analysts. They imagine themselves to focus on the environment rather than on internal mental states because the environment enters into observable relations with behavior (whereas hypothetical mental states do not) or because the environment is directly manipulable (whereas hypothetical mental states are not). Nevertheless, I believe that this type of justification is the wrong one; environment-based explanations of behavior and an inductivist and pragmatic philosophy of science are best kept apart.

The simplest way to show the kind of scientific strategy described by Chiesa (1994) does not provide a proper justification for the environmentalism of behavior analysis is to examine what would happen if this strategy were brought to bear on other scientific fields such as physics or evolutionary biology. (Remember that philosophy-of-science considerations are topic-neutral and can be applied to any scientific discipline.) The result, of course, would be disastrous. As Killeen (1984, pp. 28–29) has eloquently argued, building on observed regularities without ever invoking hypothetical constructs would have impeded the formulation of some of the most successful theories known in physics, chemistry, or biology.

The example of biology is particularly interesting in this respect because of the frequent references to the theory of natural selection in the behavior-analytic literature (e.g., Baum, 1994). When Darwin (1859) was developing his theory of evolution, he was not proceeding as an inductivist. On the contrary, he was contrasting hypothetical scenarios (for example, special creation versus descent with modification) on the basis of their predictions about current biogeographical facts. This puts Darwin squarely in the hypothetico-deductive (Ghiselin, 1969; Penny, 2009) or retroductive (McMullin, 1984) tradition.²

² Ghiselin (1969) and Penny (2009) speak of the “hypothetico-deductive method” or “hypothetico-deductive model” with respect to Darwin. In general I prefer to avoid these terms, lest the reader infer that anyone who uses retroductive explanations endorses all aspects of Popper’s (1963) philosophy of science. In fact proponents of the retroductive strategy can and do criticize Popper’s falsificationism (even though they invoke hypothetical constructs in their explanations), for example on the ground that falsificationism aims at refuting one conjecture at a time instead of confronting multiple hypotheses simultaneously (e.g., Hilborn & Mangel, 1997). Along similar lines, let me emphasize that adherence to an inductivist philosophy of science of the sort described by Chiesa (1994) or Skinner (1950) entails no commitment to the feasibility of inductive, as opposed to deductive, logic (cf. Musgrave, 2011). Also see Footnote 1.
TONNEAU

The behavior analyst’s preference for environmental explanations over a scientific analysis of internal mechanisms would be similarly disastrous in contexts other than psychology. Applied to physiology, for example, the environmentalism of behavior analysis would have impeded not only the formulation of scientific accounts that are basically correct, but the development of successful treatments that depend on such scientific accounts—leaving us at the mercy of holistic medicine, as Killeen (1984, p. 28) quipped. That there are no general justifications of the philosophy-of-science type for concentrating on environmental histories as an explanatory resource shows that the proper justification, if any, for explanations that are both environmental and historical must be indigenous to psychology.

What would make it not only correct and important, but also essential, in psychology to propose explanations of behavior in terms of an animal’s past interactions with its surroundings? Two conditions are needed, one empirical and one conceptual. The empirical condition would be that organisms, at least those of interest to psychologists, are indeed sensitive to such histories—and not just unstructured histories, but histories that include patterns of instantiations of complex relations among events. A century of experimental research on “memory” (Wilcox & Katz, 1981), to say nothing of daily life and clinical cases, shows that this condition holds. Not only are animals and people sensitive to structured environmental histories, but which portions of these histories guide behavior from moment to moment depends on the local context and its own relations to previous events (e.g., Tonneau, 2011, 2013). By contrast, the entities that compose the fields of physics and chemistry, even dynamical systems, show little sensitivity to structured histories, and no shifts of behavioral guidance from one instant to the next between different portions of these histories.

Now the behavioral effects of environmental histories, however complex and shifting, are mediated causally by physiological, largely neural, mechanisms. What is wrong with explaining behavior in terms of these internal mechanisms instead of extended histories? From a philosophy-of-science perspective, nothing; and we know that suggesting otherwise leads to unwelcome consequences (see above). But if we assume that interactions with the environment across time are constituents of psychological phenomena, physiological explanations of behavior, however adequate scientifically, will not qualify as psychological explanations. This assumption is the second, conceptual, condition for focusing on past environmental interactions instead of internal mediation, and it makes historical explanations essential to psychology. Notice that this second condition, like the first, is specific to psychological phenomena and unrelated to philosophy of science, Ryle’s (1949) and Rachlin’s (1994) views of the mental, for example, or Kantor’s (1985, p. 5) definition of psychological behavior as differentiable and modifiable across time, make it intrinsic to explanations of behavior that they must be historical if they are to be psychological at all. And indeed, a philosopher of mind such as Ryle (1949) is commonly (and correctly) recognized as a behaviorist, even though he never cared about the scientific prediction and control of behavior.

If the proper justification for an environment-based explanation concerns the nature of psychological phenomena and nothing else, then the behavior analysts who reject cognitivism or mentalism have little reason to stick to an inductivist, Humean philosophy of science. Not only does this philosophy fail with respect to a
NON-HUMEAN BEHAVIOR ANALYSIS

variety of scientific explanations outside of psychology (Killeen, 1984), it also impedes, as I will now argue, the development of environment-based psychology. The restrictions that come with an inductivist, Humean philosophy of science make behavioristic, environment-based accounts of action less successful than they could be. To understand why, we must reexamine the issue of theory in behavior analysis.

Theoretical Properties in Behavior Analysis

Whether radical behaviorism allows behavior analysis to be a theoretical discipline is difficult to judge because any discussion of theoreticity (what it means of something that it is theoretical) raises complex questions in philosophy of science. The issue is complicated further by the fact, noted by Burgos (2007), that no matter how non-theoretical the radical behaviorist’s scientific strategy appears to others, radical behaviorists insist that it is nevertheless theoretical (e.g., Skinner, 1969, p. viii). Of course, anyone is free to define “theoretical” the way he or she pleases, with the result that one’s own scientific strategy can always come out as “theoretical” with enough semantic massaging.

I propose to cut through these difficulties in three simple steps. First, I define a criterion, or filter, that a property must satisfy to count as theoretical with respect to a data domain. Second, I show that this criterion excludes all of the properties that are acceptable in the radical behaviorist philosophy of science. Finally, I give a concrete example of a property that does satisfy this criterion (and is therefore unavailable in a radical behaviorist framework) and has proved useful in furthering the prediction and control of operant behavior.

Let us start with the concept of a data domain, more specifically, the concept of behavioral data domain. Along the full course of its life from birth to death, any organism encounters a long sequence E of environmental events and exhibits a long sequence B of behaviors. Together, these two sequences, E and B, may be said to cover the organism’s life. Call any <E, B> pair that could possibly cover an organism’s life, a variant of the behavioral data domain, and call the set of such variants, the behavioral data domain. The delineation of this domain obviously depends on what counts as environmental or behavioral. Fortunately there is a good deal of agreement on this issue (even though few may agree on the definition of “behavior”). That some cases (e.g., neural firings: are they behavioral?) are controversial does not matter. Just throw the controversial cases away and keep the common core.

Now consider a property Q that could assume different value (for example, different magnitudes). I will say that Q is fixed by a data domain just in case for each variant D of this domain and for each part of D, either (a) this part does not instantiate Q, and cannot instantiate it, or (b) this part instantiates Q with value q, and cannot fail to instantiate Q with value q.³ Put more intuitively, when Q is fixed

³ If Q cannot assume different values, the definition of fixing is simpler: Q is fixed by a data domain just in case for each variant of this domain and each part of this variant, either (a) this part does not instantiate Q, and could not instantiate it, or (b) this part instantiates Q, and could not fail to instantiate it. In either definition, a “part” is taken
by a data domain, the composition of the latter fully determines whether $Q$ is instantiated and if so, with what value. The notion of a property being fixed by a data domain, relying as it does on metaphysical necessity (“cannot” or “cannot fail”), is closely related to property entailment and supervenience in analytic philosophy (e.g., Kim, 1984). In fact, property fixing is a case of supervenience (Kim, 1988), but “fixing” makes the discussion of technical issues more intuitive and easier to follow (see below).

Finally, let me claim that any property $Q$ that is fixed by a data domain is not theoretical relative to it; equivalently, any quantity $Q$ that is theoretical with respect to a data domain is not fixed by it. This claim is intended to capture an important truth about the properties that are widely recognized as theoretical in biology, chemistry, or physics: whether they are instantiated or not, and if so, with what values, is not determined by the corresponding data domain.

This criterion of theoreticity is the mirror image of one introduced by Sneed (1971) in The Logical Structure of Mathematical Physics (pp. 31-35). Put intuitively, according to Sneed, a property $Q$ is theoretical with respect to some theory $T$ just in case we need to assume the truth of $T$ (in the form of various hypotheses about the measurement process) to be able to measure $Q$. Neither Sneed nor I appeal to unobservability in defining theoreticity. Some theoretical properties may well be unobservable (whatever that means), but they need not be so; theoreticity and unobservability are different issues. Also, both Sneed and I agree that there is no such thing as theoreticity in general. What is theoretical in one context may be non-theoretical in another. Sneed’s concept of theoreticity is relative to a theory $T$, however, whereas mine is relative to a data domain $D$. Yet the criteria are compatible. On the assumption that theories are underdetermined by the data to which they are applied, a property theoretical with respect to a theory $T$ (and the measurement of which, therefore, presupposes $T$) is not fixed by the corresponding data domain.

The main difference between Sneed’s (1971) criterion and mine is that the former gives necessary and sufficient conditions for theoreticity relative to a theory $T$, whereas my criterion (or filter) gives only a necessary, not sufficient, condition for theoreticity relative to a data domain $D$. A property may well pass my criterion and still fail to qualify as theoretical, for example because the relations in which it participates are not general enough to support explanatory unification (e.g., Kitcher, 1981). Be that as it may, the criterion I propose for theoreticity with respect to the behavioral domain has important implications for behavior analysis.

Properties and Fixing

Take any property $Q$ that we commonly think of as theoretical with respect to a data domain—for example, force magnitude with respect to kinematics. Why is this property left unfixed by the data domain? Because the measurements made in the latter arise from causal processes that involve not only $Q$, but other entities and properties as well. Often the theory under evaluation comprises (or is conjoined to be any (connected or disconnected) portion of space-time.
NON-HUMEAN BEHAVIOR ANALYSIS

with a specification of how these interact with $Q$ to determine the measurements made in the data domain. Classical mechanics, for example, includes force composition laws that state how different forces combine to produce acceleration in any inertial frame. Whereas force composition in classical mechanics is deterministic, it is common to assume that the measurements performed in the data domain arise from $Q$ together with random factors $\varepsilon$ for which only a probability distribution is specified.

In some cases, we can devise a computational procedure (say, ordinary least-square regression) that delivers some unique value $Q^*$ as an estimate of $Q$. In these conditions the estimator, $Q^*$, is fixed by the data domain. But the theoretical quantity of interest, $Q$, is not. In fact it could assume any value whatsoever, depending on how the random factors ($\varepsilon$) combine with $Q$ to determine the values of the properties that are measured. In other cases (such as nonlinear regression), the estimation procedure for $Q$ is not even guaranteed to deliver a unique value. This makes it even more evident that theoretical quantities are not fixed by the corresponding data domain.

The relevance of the concept of fixing to radical behaviorism as a philosophy of science is that the latter makes no room for theoretical properties (in the current sense) with respect to the behavioral data domain. As described by Chiesa (1994), radical behaviorism starts with correlations between environment and behavior and then builds upward with higher-order variables extracted from lower-order correlations. All of these, however, are fixed by the behavioral data domain. Any $<E, B>$ pair of environmental and behavioral sequences that could cover an organism’s life, that is, any variant of the behavioral data domain, fixes the value of any correlation between environmental and behavioral events within this $<E, B>$ pair as well as the value of any higher-order variable extracted from them. Remember that these variables are not estimates of other quantities—rather, they are the very variables on which the inductive strategy of radical behaviorism is built. The additional inductive step of passing from a finite set of particulars to universal quantification does not introduce any new properties. All of them are fixed by the behavioral data domain and remain so.

My conclusion about the status of theoretical quantities in behavior analysis contrasts with those of a classic paper by Shimp (1976) on memory and behavior. In his article, Shimp discusses a measure of sequential dependency that Tulving (1962, 1964) used to analyze the results of free-recall studies. This measure is:

$$\sum \frac{n_{ij} \log(n_{ij})}{\sum n_i \log(n_i)},$$

(1)

where $n_{ij}$ is the number of times the $(i, j)$ word pair appears in the subject’s output, $n_i$ being the output frequency for word $i$. Shimp states that (1) is an “unobservable theoretical quantity” on the ground that it is “calculated from data according to a formula derived from a theory” (p. 116), in this case information theory. Other examples of unobservable theoretical quantities according to Shimp are absolute response rate (p. 120) and relative response rate on a concurrent schedule (p. 121).
Branch (1977) has objected to Shimp (1976) that response rate as well as the measure used by Tulving “are derived directly by mathematical operations from observables. No inferential leaps are made, and most observers would not refer to either as theoretical constructs” (p. 173). The core of Branch’s criticism is correct. Admittedly, Tulving would never have used (1) as a measure of organization if he hadn’t known about information theory. Furthermore, Shannon (1948) did derive a formula analogous to (1) from axioms that any measure of organization “should” intuitively satisfy (continuity, partitioning, etc.). Shannon’s derivation, however, did not involve any hypothesis about the extra-mathematical world. His derivation was no different in kind than the construction of an integral from basic desiderata about what the measure of the size of a set “should” be. Furthermore, once (1) has been derived mathematically as a quantity that satisfies our desiderata, we are free to employ it or not as a measure of organization. If we do employ it, however, (1) functions as a definition of the chosen measure, not as a scientific hypothesis about its values—a scientific hypothesis that could conceivably be false. Because (1) defines our measure, the latter is fixed by the behavioral data domain (through the \( n_i \) and \( n_{ij} \) word counts) and cannot be theoretical with respect to it.

Of course, (1) may well be taken to estimate another property, for example a cognitive property that would be theoretical relative to the behavioral data domain. Tulving (1962) himself was well aware of the distinction between (1) and the property it was supposed to estimate:

*Behavioral manifestations* [italics mine] of the hypothesized organizing process can best be studied under conditions where the order in which the subject recalls items is free to vary … a quantitative analysis of sequential dependencies among items in subject’s free recall on successive trials would yield measures of a response variable **closely related** [italics mine] to the hypothetical organizing process. (p. 345)

That the quantity \( Q^* \) defined by (1) was merely an estimator of some theoretical property \( Q \) “closely related” to \( Q^* \), but nevertheless distinct from it, is clear from this quotation. Admittedly, Tulving referred to (1) as a measure of “subjective organization” (p. 345), but only because the degree of behavioral organization \( Q^* \) measured by (1) was supposed to arise from the subject’s own cognitive structure (\( Q \)) instead of being constrained by the verbal material. Like response rate or any other quantity admissible in the radical behaviorist philosophy of science, (1) is non-theoretical relative to the behavioral data domain.

**The Case of Reinforcement “Value”**

Introducing properties that are theoretical with respect to the behavioral data domain, however, has proved effective in furthering the scientific understanding of behavior. Consider the concept of *reinforcement value* (e.g., Commons, Mazur, Nevin, & Rachlin, 1987). Mazur (1987) applied this concept to a schedule that let the animal choose between two options: a standard option with amount of food \( A_s \) delivered after \( t_s \) seconds, and an adjusting option with amount of food \( A_a \) delivered after \( t_a \) seconds. Whereas the delay of the standard option \( (t_s) \) stayed constant, the...
delay of the adjusting option \( t_a \) was increased whenever the animal chose this option and decreased whenever the animal chose the fixed option. The result of this adjusting procedure was that the animal eventually showed indifference between the two options, choosing either one on about 50% of the trials.

Mazur (1987) argued that at indifference, the value of the adjusting option, \( V_a \), equalled the value of the standard option, \( V_s \). He also tested various hypotheses about how the delay \( t_i \) of option \( i \) affected its reinforcement value \( V_i \). The following assumption:

\[
V_i(t_i) = \frac{Q_i}{1+kt_i}
\]  

(2)

with \( Q_i = V_i(0) \), predicts that at indifference, the delay for the adjusting option will be a linear function of the delay for the fixed option:

\[
t_a = \left( \frac{Q_a}{Q_s} \right) t_s + \frac{1}{k} \left( \frac{Q_a}{Q_s} - 1 \right).
\]  

(3)

This prediction was tested by manipulating \( t_i \) across phases and noting that \( t_a \) was indeed a linear function of \( t_i \) (Mazur, 1987), as implied by Equation (3). Other hypotheses about \( V \) would have predicted a different relation between \( t_a \) and \( t_s \).

This approach was later extended to schedules that presented food after a fixed number of responses. Mazur and Kralik (1990) argued that any event, be it food or a response, had its instantaneous reinforcement value \( Q_i \) modified by the delay \( t_i \) since the initial choice in accordance with Equation (2). The overall value of an event of duration \( d_i \) starting \( t_i \) seconds after the initial choice was taken to be the integral of Equation (2) from \( t_i \) to \( t_i + d_i \):

\[
V_i(t_i, d_i) = \int_{t_i}^{t_i+d_i} \left( \frac{Q_i}{1+kx} \right) dx = \frac{Q_i}{k} \ln \left( 1 + \frac{k d_i}{1 + k t_i} \right).
\]  

(4)

Again, Equation (4) gave a good account of the indifference functions (curvilinear, this time) between schedules that required different numbers of responses and presented food at different delays (Mazur & Kralik, 1990).

The research strategy behind Equation (2) is the exact opposite of the radical behaviorist philosophy of science. Research on reinforcement value did not start by collecting data from the adjusting schedules, noticing a linear relation between the adjusting delay and the standard delay, and then using the best-fitting slope between \( t_a \) and \( t_s \) as a higher-order dependent variable. Instead, the linear relation was predicted from a particular hypothesis about reinforcement value, and the adjusting schedules were devised so as to test this hypothesis (see Mazur, 1987).
Also, the theoretical property of reinforcement value is not fixed by the behavioral data domain. Although I have written Equation (3) in a simple deterministic form, it should actually include an error component with an associated probability distribution. And even if the data followed Equation (3) perfectly, it would still be possible for reinforcement value to assume another form than Equation (2), depending on which auxiliary hypotheses (such as the supposed equality of $V_a$ and $V_s$ at indifference) are brought to bear on the behavioral data domain.

How can the radical behaviorist interpret a hypothetical construct such as reinforcement value? One suggestion, due to Chiesa (1992, pp. 1293-1297), is to employ this type of construct as a provisional help in the discovery of functional relations, and then discard it once the latter have been established. The problem with this strategy is that getting rid of the construct often gets rid of explanatory unification as well. Thus, the indifference functions observed in the laboratory appear linear (Mazur, 1987) or curvilinear (Mazur & Kralik, 1990) depending on what schedule is used. Whereas the inclusion of a Riemann integral in Equation (4) gives an explanation to these appearances, discarding reinforcement value would leave us only with a catalog of isolated functions. Skinner (1950) did recognize that catalogs are not enough and that “a formal representation of the data reduced to a minimal number of terms” (p. 216) is needed. The terms that support this unification, however, do not arise inductively from functional relations (p. 216). In the case of reinforcement value at least, the theoretical term comes before the functional relations (e.g., Equation 3) and unifies them at a deeper level than the behavioral data domain.

Because radical behaviorism has no room for reinforcement value, the operant researchers who rely on this concept often give it a cognitive flavor, calling it, “subjective value” (e.g., Green & Myerson, 2004, p. 769). An alternative to this cognitive interpretation is instrumentalism—that is, to employ the concept of reinforcement value without giving it any ontological significance (e.g., Rachlin, 2005). After admitting that “in cognitive theory the value of an alternative is the state of an internal process” (p. 157), Rachlin (1989) states:

> In behavioral theory, value is a stage of a process that the observer goes through rather than the subject. It is a sort of accounting method used to summarize the results of previous observations in a form that can be easily used to predict future observations. (p. 158)

Rachlin relates this perspective to operationism (p. 167). Instrumentalism with respect to reinforcement value can also be mixed with cognitivism, as when Logue (1988) interprets Equation 2 to mean that behavior depends on the “perceived” (that is, subjective) rather than physical, properties of reinforcers (p. 679), but adds that cognitive stages of processing are postulated merely for their heuristic merit and “do not necessarily exist” (p. 702).
NON-HUMEAN BEHAVIOR ANALYSIS

Invisible Causation

I now show that properties such as reinforcement value, which are theoretical relative to the behavioral data domain, can be given a realist interpretation that is fully behavioristic but neither cognitive nor physiological. This interpretation is made possible by endorsing the environmentalism of radical behaviorism while rejecting its Humean philosophy of science. As Chiesa (1992, p. 1289) pointed out, Humean views on causation\(^4\) have been, and remain, extremely influential in philosophy. Their influence has been eroding little by little, however (e.g., Cartwright, 1981; Scriven, 1971; Tooley, 1993), due in no small measure to the incapacity of Humean philosophers to come up with a satisfactory analysis of causality in non-causal terms (Brand, 1979). So far, all reductive programs with respect to causation have failed (Koons, 2000).

In this context, causal primitivism or anti-reductionism (Carroll, 2012) is an increasingly attractive philosophical position. It assumes that causation is a primitive, basic constituent of the world, of the sort that cannot be reduced to any non-causal Humean regularity or non-causal physical relation. Assuming that we observe two states of affairs A and B together, whether A causes B is tested by removing A and noting whether B persists or not. Any causal relation that exists between A and B, however, is supposed to exist regardless of whether testing is carried out. Besides, the test is fallible, no matter how many times it is repeated. The AB correlations that hold across tests might be coincidental (however unlikely). Causal relations, being theoretical entities hidden behind the empirical symptoms (Armstrong, 2004), are never fixed by Humean correlations. That empirical regularities can only provide estimators ($\hat{Q}$) for causal quantities ($\bar{Q}$) is even more obvious when an effect arises from multiple sources and the associated causal interaction equation includes random factors (see below).

I further assume that causation is invisible in the sense of being absolutely unobservable. There may be no observational difference, for instance, between an AB pairing that is causal and one that is coincidental. I assume that this lack of observational difference persists down to the microphysical level. No matter which technology we use, opening up an entity and examining its components will never reveal anything more than empirical correlations (the invisibility of causation being one of the facts about which I think Hume was right). Primitive causation is nevertheless supposed to infuse reality at all levels of organization down to

\(^4\) Whether the views on causation that are commonly attributed to Hume were actually his is a matter of scholarly debate (e.g., Blackburn, 1990). Be that as it may, I use the label “Humean” to designate regularity theories of causation as well as the positivist program of replacing causation by “functional relations” (e.g., Skinner, 1953, p. 23). Similarly, I refer to “Skinnerian inductivism” even though some of Skinner’s own ideas (for example, the reflex reserve: Skinner, 1938) were actually inconsistent with it. In either case, a label is used to identify an influential viewpoint regardless of historical correctness. What matters to me are the philosophical proposals and their current influence, not who held them and with what consistency.
fundamental physics (Esfeld, 2010). Composition relations between material objects and their parts are themselves assumed to be causal. In animals with a central nervous system, for example, the causal relations between environment and behavior are just as real as (and are in fact caused by!) causal chains from the environment to the brain and from the brain to behavior (Tonneau, 2011). The causal influence of the environment on behavior is in no danger of vanishing after we find out which physiological mechanisms enable it to occur.

When an environmental property \( x \) influences a behavioral property \( y \) and both are continuous variables, the magnitude of \( y \) will depend on the magnitude of \( x \) and on how much \( x \) influences \( y \). Call the latter factor the degree of influence of \( x \) on \( y \). The simplest hypothesis about the relation of \( y \) to \( x \) is:

\[
y = wx,
\]

where \( w \) is the degree of influence of \( x \) onto \( y \). Equation 5 is incorrect as it stands, however, because strictly speaking \( y \) always depends on factors other than \( x \) (even though the latter may be negligible). It is commonly assumed that these factors combine additively and that their sum, \( \varepsilon \), is Gaussian. In these conditions, Equation 5 should be replaced by:

\[
y = wx + \varepsilon.
\]

When \( y \) depends on multiple environmental causes \( x_1, x_2, \ldots, x_n \), we are led to the formulation of a causal composition equation in which \( y \) depends not only on the magnitudes \( x_i \) but also on their degree of influence (\( w_i \)) on \( y \). Classically:

\[
y = \sum_{i=1}^{n} w_i x_i + \varepsilon.
\]

Nothing requires the composition of causal influences to be linear, however, and it is common to include a non-linear function \( f \) in the composition equation:

\[
y = f\left(\sum_{i=1}^{n} w_i x_i\right) + \varepsilon.
\]
NON-HUMEAN BEHAVIOR ANALYSIS

Equation 8 is analogous to force composition in classical mechanics. Obviously, the formulation of an equation of this type depends on a realist stance toward the individual degrees of influence $w_i$ (Creary, 1981). In general, the description of multiple causation requires assumptions about which causal influences are present and how they combine to determine their effect (T. L. Smith, 1994, pp. 112-116).

Finally, all of the $w_i$'s, being genuine aspects of the causal relation between environment and behavior, can themselves be modulated by other variables. This leads to the formulation of hierarchical causal equations in which each $w_i$ depends on environmental factors according to higher-order degrees of influence. Imagine for example a concurrent schedule with two keys L and R and in which choice is influenced by the rate ($R_L$ and $R_R$), amount ($A_L$ and $A_R$), and immediacy ($I_L$ and $I_R$) of food associated with each key. The following expression:

$$y = f_1(R_L) \cdot f_2(A_L) \cdot f_3(I_L) + \epsilon,$$

where $y$ being the ratio of L and R response counts, can be seen as a causal combination equation in which degrees of influence are different functions of rate, magnitude, and immediacy, and in which causal influences combine multiplicatively on each key to determine behavior (cf. Baum & Rachlin, 1969).

In line with the interpretation of reinforcement “value” as a subjective or internal phenomenon (Logue, 1988), Killeen (1972, p. 490) has pointed out that functions of the type that appear in Equation 9 can be thought of as subjective scales. Causal primitivism provides an alternative, however. According to causal primitivism, these functions are neither subjective nor internal. Instead, they describe the relations between environmental quantities and invisible degrees of influence, and the equation as a whole is a causal composition rule (force composition in classical mechanics being another example). A comment by Herrnstein (1990) provides an illuminating comparison of the cognitive and causal-primitivist perspectives in relation to utility:

“We may observe behavior and its objective consequences, but, however carefully we observe or deeply we probe, utility itself will remain out of reach of direct observation. The invisibility of utility is a particular instance of the inscrutability of subjective states in general. Like force in physics, utility names a relationship between observables, but is not itself observable. (p. 221)

It is true that utility (or reinforcement) is ultimately as inscrutable as force, but why? The causal-realist answer: Because reinforcement value and utility, which Herrnstein takes to be “subjective states,” actually are features of causal relations between environment and behavior, just as forces are species of physical causation (Bigelow, Ellis, & Pargetter, 1988).

Although strong realism with respect to causation is inconsistent with a Humean philosophy of science (Boyd, 1981), some behavior analysts already

25
interpret reinforcement value in a realist and causal fashion. Mazur (2001) states for example that “value refers to the strength or effectiveness of a reinforcer” (p. 97), and Shull and Spear (1987) write:

The fundamental effect of reinforcement is assumed to be an increase in the emission rate of a response. Then if reinforcer A should maintain a higher rate of response than reinforcer B, other things equal, reinforcer A would be doing more strongly [emphasis mine] what reinforcers do. (p. 187)

Here “doing” is not an issue of mere correlation or regularity but of exerting a genuine causal influence on behavior—the sort of “push” that Skinner (1938, p. 21) wanted to exorcise from behavior analysis. In contrast to Humean correlations, causal influences are invisible and must be studied inferentially.

From the perspective of causal primitivism, it is not surprising that research on reinforcement “value” (Equation 2) displays formal features analogous to the study of causal composition in other scientific fields. Mazur’s (1987) hypothesis that at indifference on a concurrent schedule, the “value” (read: causal influence) must be the same on each side is a typical causal argument. It is no different in kind from the physicist’s argument that a balance at equilibrium must have identical forces on each arm. Similarly, Mazur and Kralik’s (1990) computation of an overall degree of influence as the integral of its components (Equation 4) is equivalent to assuming the additivity of individual causal contributions. This sort of linearity assumption is commonplace in physics (e.g., Schwartz, 1972/1987), although it is often an approximation and may then be replaced by more complex interaction rules.

Conclusion

I believe that environment-based behavior science and the philosophy of science which underlies radical behaviorism should be dissociated. I see no reason why a Humean view of causation and a strictly inductive research strategy should be essential parts of environment-based psychological theory. On the contrary, I have argued that the philosophy-of-science components of radical behaviorism are deficient and unnecessarily restrict the theoretical power of behavior analysis, especially in relation to quantitative operant research.

Behavior analysts seem to have taken it for granted that environment-based explanations must be of the Humean, inductive kind (e.g., Chiesa, 1994). Causal primitivism, however, provides a philosophical rationale for endorsing constructs that are theoretical relative to the behavioral data domain yet are neither cognitive nor physiological. These constructs consist of invisible causal relations between environment and behavior. Although invisible, these relations are instantiated where their bearers are present, that is, in environment-behavior pairs. Therefore they are compatible with environment-based psychology. In particular, they are not inside the head. Nothing in the retroductive explanatory strategy requires it to deal with microscopic entities or to involve mechanisms internal to the system whose behavior we want to understand (McMullin, 1978, p. 139).

Causal primitivism allows behavior analysts to make sense of constructs such
NON-HUMEAN BEHAVIOR ANALYSIS

as “value” (e.g., Rachlin, 1989) in a straightforward fashion. Reinforcement “value” is a degree of influence, the extent to which a consequence reinforces behavior. Here reinforcement is conceived as a causal relation (Tonneau, 2008), not a set of Humean regularities. Stimulus control and elicitation also involve causal relations between environment and behavior, and are just as inferential as operant reinforcement. That the latter is an inference from the behavioral domain has been recognized by Sidman (2010), but he seems to believe that the theoreticity of reinforcement has to do with the fact that multiple observations are needed to assess its presence (p. 180). This cannot be right. I need multiple observations to evaluate the average temperature $t^*$ in my living room, but computing $t^*$ requires no inference at all (cf. Branch, 1977). Stimulus control and reinforcement are theoretical relative to the behavioral domain not because they involve multiple instances but because they involve non-Humean causal relations.

Hineline (1981) has stated that elicitation is a “theoretical construct that we accept as defined and operating at the level of behavior” (p. 162). From the perspective of causal primitivism applied to behavior, this statement is correct. Elicitation is a theoretical construct that operates at the level of behavior and nowhere else. Elicitation operates neither in the fictitious world of cognitive processes nor in the real domain of neurophysiology—although the latter gives us the internal mechanisms without which there would be no elicitation. But Hineline’s claim is incompatible with the Humean philosophy of science known as radical behaviorism. How could elicitation be a theoretical construct if it were nothing more than a correlation (Skinner, 1931)?

References


NON-HUMEAN BEHAVIOR ANALYSIS


NON-HUMEAN BEHAVIOR ANALYSIS


Author's Note
This article is based on a talk presented at the 1st Amazonian Simposium on Behavior Research and Theory (Universidade Federal do Pará, Belém, Brasil, November 2013). Jay Moore served as Action Editor for this manuscript. Correspondence should be sent to the author at Universidade Federal do Pará, Núcleo de Teoria e Pesquisa do Comportamento, Rua Augusto Corrêa, 01 Guamá, Belém, PA 66075-110, BRASIL. E-mail: francois.tonneau@gmail.com