Perspectives on Behavior Science (2020) 43:361–385
https://doi.org/10.1007/s40614-020-00256-w

ORIGINAL RESEARCH

The Study of Rule-Governed Behavior and Derived Stimulus Relations: Bridging the Gap

Colin Harte[1] · Dermot Barnes-Holmes[1,2] · Yvonne Barnes-Holmes[1] · Ama Kissi[1]

Published online: 21 May 2020

Abstract

The concept of rule-governed behavior or instructional control has been widely recognized for many decades within the behavior-analytic literature. It has also been argued that the human capacity to formulate and follow increasingly complex rules may undermine sensitivity to direct contingencies of reinforcement, and that excessive reliance upon rules may be an important variable in human psychological suffering. Although the concept of rules would appear to have been relatively useful within behavior analysis, it seems wise from time to time to reflect upon the utility of even well-established concepts within a scientific discipline. Doing so may be particularly important if it begins to emerge that the existing concept does not readily orient researchers toward potentially important variables associated with that very concept. The primary purpose of this article is to engage in this reflection. In particular, we will focus on the link that has been made between rule-governed behavior and derived relational responding, and consider the extent to which it might be useful to supplement talk of rules or instructions with terms that refer to the dynamics of derived relational responding.

Keywords Rule-governed behavior · Derived relations · Relational networks · Relational frame theory · HDML framework

The concept of rule-governed behavior or instructional control has been widely acknowledged for many decades within the behavior-analytic literature. The concept was originally proposed by B. F. Skinner in 1966 in an attempt to explain problem-solving behaviors. In particular, Skinner defined a rule as a stimulus or stimuli that specified

This article was prepared with the support of an Odysseus Group 1 grant awarded to the second author by the Flanders Science Foundation (FWO).

✉ Colin Harte
Colin.Harte@UGent.be

[1] Department of Experimental, Clinical, and Health Psychology, Ghent University, Henri Dunantlaan, 2, 9000 Ghent, Belgium

[2] School of Psychology, University of Ulster, Coleraine, UK

reinforcement contingencies, which allowed the listener to solve problems without directly contacting environmental contingencies. For example, the rule "boiling an egg takes longer at higher altitudes" allows the listener to adjust the boiling time without undercooking multiple eggs. When experimental research began to identify important interspecies differences between the performances of humans and nonhuman animals on schedules of reinforcement (e.g., Bentall, Lowe, & Beasty, 1985; Weiner, 1969), the human ability to engage in rule-governed behavior was typically used to explain this difference. The basic argument was that the human propensity to engage in rule-governed behavior undermined the sensitivity typically observed with nonhuman behavior on schedules of reinforcement. Research on rule-governed behavior also extended beyond the sensitivity issue when studies began to explore the impact of different types of rules (i.e., instructions that simply described a reinforcement schedule versus instructions on how to perform on that schedule).

The concept of rule-governed behavior has frequently been linked to the distinct nature of human psychological suffering. In particular, it has been argued that excessive reliance upon rules and rule-following at the expense of contingency-sensitive behavior may, for example, cause an individual to miss out on many opportunities for reinforcement in the natural environment, which is associated with depression (e.g., Abramson, Seligman, & Teasdale, 1978; Hayes, Strosahl, & Wilson, 1999; Seligman, 1974). The concept of rules, or instructional control (note that we will use these terms interchangeably throughout), has also been linked to equivalence class formation (see Sidman, 1994) and to derived relations in general (see Hayes, 1989).[1] For example, it has been argued that rules specify contingencies because some of the words in a rule or instruction participate in equivalence relations with some of the stimuli or events to which the rule refers (Hayes & Hayes, 1989).

In general, the concept of instructions or rules would appear to have been relatively useful within behavior analysis and few would seriously question this conclusion. However, it seems wise from time to time to reflect upon the utility of even well-established concepts within a scientific discipline. Doing so may be particularly important if it begins to emerge that the existing concept does not readily orient researchers toward potentially important variables associated with that very concept. The primary purpose of this article, therefore, is to engage in that very reflection (see Dixon, Belisle, Rehfeldt, & Root, 2018, for a broadly similar reflective exercise). In particular, we will focus on the link that has been made between rule-governed behavior and derived relational responding, and consider the extent to which it might be useful to extend the talk of rules or instructions with terms that refer to the dynamics of derived relational responding. As explained later in the article, we use the term "dynamics" to refer to the ways in which units of experimental analysis may interact with each other.

[1] The literature on stimulus equivalence and derived relations in general is focused on reinforcing or training specific subsets of relational responses such as matching A to B and B to C, and then observing the spontaneous emergence of untrained responses such as matching A–C and C–A. Other untrained responses may also emerge when a specific function is trained to a stimulus participating in a derived relation (e.g., if A, B, and C participate in an equivalence relation, and A is paired with a reinforcer, stimulus C may then acquire reinforcing functions in the absence of direct pairing). This latter effect has often been referred to as a derived transformation of functions.

The article will begin with a brief history of the behavior-analytic literature on rule-governed behavior. It will then consider how the concept of derived stimulus relations has allowed for a more precise experimental analysis of human verbal behavior in general, and, in more recent research, instructional control in particular. We will then consider how recent conceptual and empirical developments could be used to advance research in the area of rule-governed behavior, and examine examples of empirical studies in which this has already begun to emerge. Finally, we will argue that, in light of these developments, it may be useful to discuss behavior traditionally referred to as rule-governed in terms of the dynamics of derived relational responding.

A Brief History

The Early Years: Instructions and Schedule Insensitivity

As mentioned above, rule-governed behavior as a concept was first introduced by Skinner in 1966 and rules were defined as contingency-specifying stimuli. This definition of rule governance was frequently invoked when attempting to explain human performance on schedules of reinforcement, which appeared to differ from the schedule-induced patterns of responding typically observed with nonhuman organisms. For example, both Leander, Lippman, and Meyer (1968) and Lippman and Meyer (1967) gave participants minimal instructions for responding on a fixed interval (FI) schedule (i.e., instructions that did *not* specify the temporal nature of the schedule). The researchers found that two different patterns frequently emerged: a high-rate pattern in which responding was maintained at a high rate more or less throughout each interval, and a low-rate pattern in which participants emitted only a few or no responses until the end of the interval. It is important to note that these two different patterns appeared to be correlated with two different types of self-report when participants were asked about their performances at the end of the experiment. In particular, those individuals who had produced high-rate patterns frequently reported that reinforcers were delivered based on the number of responses emitted, whereas those who produced low-rate patterns typically reported that reinforcers were delivered only after a particular amount of time had elapsed. Based on these initial findings, the researchers suggested that human schedule performances were influenced by the types of rules participants generated during exposure to the schedules, rather than through direct contact with the scheduled contingencies per se.

Related research by Weiner (1964, 1969) involved giving instructions to participants that sought to establish these two different types of performance on FI schedules. In particular, if participants were instructed that reinforcers were contingent on number of responses, they tended to produce high-rate patterns, but if they were instructed that the availability of reinforcers was contingent on the passage of time they tended to produce low-rate patterns. This research thus demonstrated that instructions appeared to influence the way in which participants interacted with reinforcement schedules, which served to maintain particular response patterns, even if the instruction was false. For example, if a participant was told that reinforcers were delivered contingent on large numbers of responses, and they responded at a high rate on an FI schedule, and although the rule was inaccurate it appeared to be correct because reinforcers were

indeed delivered regularly following large numbers of responses (during the fixed interval).

Additional Evidence that Rule-Governed Behavior was Involved in the Insensitivity Effect

Yet further evidence suggesting that verbal rules played a role in schedule insensitivity was obtained with infants and young children. For example, Lowe, Beasty, and Bentall (1983) found that infant performances on FI schedules closely resembled that of rats and pigeons, rather than the high and low rates that had been observed so often with adult humans. The authors argued that the nonhuman-like performances produced by the infants could be explained by their inability to formulate rules about the contingencies, and thus they interacted directly (i.e., with sensitivity) with the schedules.

Follow-up research by Bentall et al. (1985) repeated the earlier study with infants up to and beyond the age of 2½ years. They replicated the earlier findings with young infants, but found that older children up to age 5 years displayed some variations in performance. In particular, some children produced schedule performances similar to those of the young infants, whereas others produced more adult-like patterns. Children over the age of 5 tended to produce patterns that were consistently adult-like (i.e., schedule insensitive).

Further research also emerged with regard to how, under certain circumstances, it was possible to reduce the apparent impact of instructions, thus yielding more nonhuman-like performances even in verbally able adult participants. For example, some researchers gave participants tasks to perform (e.g., mental arithmetic) during exposure to FI schedules in an attempt to prevent them from "counting out" the temporal interval (e.g., Laties & Weiss, 1963). And indeed, in general these tasks disrupted participants' regular interval performances to varying degrees. Other studies provided participants with two different ways to determine whether a reinforcer was available on a particular schedule of reinforcement (Lowe, Harzem, & Bagshaw, 1978a; Lowe, Harzem, & Hughes, 1978b). Both studies by Lowe et al. involved a condition in which participants could press a panel to see if a green light was on, thus indicating that a reinforcer was available on an FI schedule; in another condition pressing the panel provided access to a digital clock that indicated how many seconds had elapsed since the last reinforcer was delivered.

In the "green-light" condition, participants tended to produce a "break-and-run" pattern of responding (i.e., an abrupt shift to a high rate of responding), but in the digital-timer condition, scalloped patterns (i.e., positively accelerated curves) were observed (i.e., the latter patterns were similar to the typical nonhuman patterns produced on FI schedules). The authors argued that the scalloped patterns were highly sensitive to the scheduled contingencies, in that they resembled the response patterns of nonhumans on FI schedules. The authors went on to argue that in the "green-light" condition participants may have relied on self-cues or rules to know when to respond, but in the "digital clock" condition participants relied far less on such rules and simply checked the timer whenever they felt they were getting closer and closer to the availability of a reinforcer. As such, the latter condition tended to produce behavior patterns that appeared to share the schedule sensitivity observed with nonhumans (who lacked the ability to generate rules or self-cues to control their schedule performances).

Analyzing Rule-Governed Behavior for Its Own Sake

Much of the early research on rule-governed behavior appeared to focus on it as the basis for explaining why (verbal) humans often produce behavior patterns that appeared to be insensitive to schedules of reinforcement, in the sense that those patterns diverged from nonhuman-like patterns. During the 1980s and 1990s, however, researchers began to focus more on the impact of rules or instructions on human schedule performances for its own sake, rather than as a way of explaining why humans and nonhumans produced discrepant patterns of responding on reinforcement schedules. Indeed, this shift away from insensitivity-focused research could be seen as a milestone in the development of the behavior-analytic study of human verbal behavior in its own right. This view was articulated towards the end of the 1980s by Hayes, Zettle, and Rosenfarb (1989):

> The "insensitivity" literature is misnamed. It is not that people become insensitive to direct experience—it is that they are relatively sensitive, under many but probably not all conditions, to control by verbal stimuli. Control by verbal stimuli is itself ultimately based on direct experience, but the experience is now remote, and it is of a special sort, resulting in stimulus control with special properties.
> The experimental task is thus not to understand the insensitivity effect per se. When that is taken as the task the literature quickly becomes trivialized. For example, papers appear that declare that people are in fact controlled by direct experience, if the direct experiences are salient enough, important enough, or held in place long enough. This may be true, but it misses the point. The value of the so-called insensitivity literature is that it provides a means of studying verbal stimuli and verbal behavior. The insensitivity preparation is a preparation, not an end in itself. (pp. 215–216)

In this context, it is worth noting that when research is less focused on the so-called insensitivity effect, human schedule performances in naturalistic contexts may approximate those performances observed with nonhumans (Hantula & Crowell, 2016; Hantula, Brockman, & Smith, 2008). Of course, similar response topographies in human and nonhuman behavior does not necessarily imply the complete absence of verbal stimulus control for humans—the role and extent of such control remains an empirical matter.

The shift from insensitivity-focused research has been marked by studies that began, for example, to compare the impact of shaping versus instructing behavior on schedules of reinforcement. For instance, Shimoff, Catania, and Matthews (1981) compared shaped performances to instructed ones on low-rate–responding schedules. In the first experiment in this study, low rates of responding were maintained by a random interval (RI) schedule with a superimposed differential-reinforcement-of-low rate (DRL) schedule. When the DRL was relaxed (that is, increased responding would lead to an increase in reinforcers), the response rate increased for most participants in the shaping group, indicating increased contact with the schedule contingencies. For the majority of participants in the instructed group, however, the response rate did not increase. In a second experiment, the researchers replaced the RI schedule with a random ratio (RR) schedule. Once again, rate of responding increased for the majority of participants in

the shaping group when the DRL was relaxed, but did not increase for the majority of the instructed participants. The authors argued that the greater insensitivity observed in the instructed group was a defining feature of rule-governed behavior.

Related to this, Catania, Matthews, and Shimoff (1982) investigated whether verbal formulations of the contingencies within a task that were shaped or instructed were in line with participant performance when the schedule contingencies were reversed. During interruptions in the schedule before the contingency reversal, one group of participants were asked to write down the performance they thought was required, whereas another group were told exactly what to write down (e.g., write "press slowly" for the left button). For those participants asked to guess, their guesses were shaped using points worth money as reinforcers. When verbal formulations (guesses) had been instructed, responding on the schedule was inconsistent among participants (i.e., responding sometimes corresponded with verbal formulations, but sometimes did not). When verbal formulations had been shaped, however, responding on the schedule was much more likely to be consistent with the shaped guesses, than with the scheduled contingencies.

Related research focused on the extent to which shaped *performance* descriptions (e.g., "press fast" for the right button) were in line with, or opposite to, shaped *contingency* descriptions (e.g., the button works "after a random number of presses"; Matthews, Catania, & Shimoff, 1985). In particular, during interruptions of the schedule, participants were asked to fill out sentence guessing sheets, as in Catania et al. (1982). The researchers reported that shaping performance descriptions produced responding that was consistent with the descriptions, rather than with the schedule. However, shaping contingency descriptions produced responding that was schedule-consistent, but inconsistent with the shaped descriptions. Overall, these studies demonstrated that it was useful to conceptualize the verbal behavior that occurs before and during exposure to a schedule as functionally distinct from the performance on the schedule itself.

In a similar vein, other researchers investigated the interaction between sources of reinforcement and instructional control (e.g., Hayes et al., 1985; Hayes & Wolf, 1984), whereas others examined the impact of different types of rules on schedule performance (Hayes, Brownstein, Haas, & Greenway, 1986). The latter study provides a compelling example of why it is important to examine the subtle interactions that may occur between instructions and schedules of reinforcement. In particular, the study highlights the *functional* properties of these interactions, rather than simply classifying a performance as sensitive versus nonsensitive. The researchers gave participants different types of instructions prior to a multiple schedule of reinforcement (Fixed Ratio [FR]-18/DRL-6) that included: (1) no instructions on the appropriate rate of responding; (2) instructions that only specified responding on one aspect of the reinforcement schedule (FR or DRL responding, but not both); or (3) accurate instructions that specified responding on both aspects of the schedule. Participants responded over two 32-min sessions (Phase 1), followed by an extinction session (Phase 2). When participants failed to demonstrate sensitive responding to the contingency changes in the first phase, this was often followed by a failure to demonstrate sensitive responding in extinction. However, when participants displayed sensitive responding to changes within the first phase, this was often followed by sensitive responding in extinction, but only if responding in Phase 1 could not be attributed to an experimenter-given, fully

accurate instruction. That is, when participants were provided either with no rule or a rule that only described part of the schedule contingencies (i.e., Points 1 and 2 above), sensitivity in Phase 1 correlated highly with sensitivity in extinction. However, when participants were provided with an accurate rule that fully described the schedule contingencies (Point 3 above), this relationship with sensitivity was not present in extinction.

The authors argued that although behavior under the influence of rules may appear topographically identical to contingency-shaped behavior at first, they may in fact be functionally distinct, as shown by the differential responding observed during extinction. Indeed, Shimoff, Matthews, and Catania (1986) also demonstrated that behavior that at first appears to be contingency-shaped may, under closer inspection, be under the control of instructions.

The type of research reviewed in this subsection highlights that there was considerable value in studying rule-governed behavior or instructional control for its own sake (i.e., because it generated an experimental analysis of human verbal behavior). Indeed, the work showed that rule-based control could be conceptualized as functionally distinct from schedule-controlled responding, even if the two classes of behavior frequently interacted with each other. This work involved distinguishing between rules that were simply provided or given to participants (by the researchers) from rules that were first generated by the participants and then shaped by the researchers; the former rules tended to produce behavior that was less sensitive to changes in schedule contingencies than the latter. The research strategy in which rule-governed behavior was separated into different types was also reflected in efforts to identify broadly defined functional classes of rules or rule-governed behavior itself, which we will consider in the next section.

Different Types of Rule-Governed Behavior: Pliance, Tracking, and Augmenting

In an attempt to identify broad functional classes of rule-governed behavior, Zettle and Hayes (1982) proposed the concepts of pliance, tracking, and augmenting. *Pliance* was defined as rule-governed behavior that is controlled mainly by consequences mediated by the speaker for correspondence between the rule and behavior (e.g., when a young child follows the rule "Eat your vegetables and I will buy you an ice cream afterwards"). *Tracking* was defined as rule-governed behavior that is controlled mainly by the correspondence between the rule and the arrangement of the natural environment (e.g., when a young child follows the rule "Eat all of your dinner and you won't be hungry later"). In the first case, the consequence (i.e., ice cream) is actually provided by the speaker and pliance as a response class thus depends on the extent to which the speaker, or functionally similar individuals, have actually mediated the delivery of reinforcers in the past. In the second case, the consequence may occur in the complete absence of any mediation by the speaker, and tracking as a response class thus depends upon the extent to which the speaker, or functionally similar individuals, have proven to be "reliable sources of information" in the past.

Finally, *augmenting* was defined as a type of rule-governed behavior that may occur in conjunction with pliance or tracking, and alters the extent to which the consequences specified by the rule have reinforcing or punishing value (Törneke, Luciano, & Valdivia-Salas, 2008). Zettle and Hayes (1982) proposed two types of augmentals,

namely *formative* and *motivative*. In the first case, for example, the statement "This token will get you a free ice cream" would be considered a formative augmental if it established the token as a reinforcing consequence. In the second case, the statement "It's really hot today, I bet you'd like an ice cream" would be considered a motivative augmental if it momentarily increased the extent to which actual ice cream functions as a potential reinforcing consequence.

The amount of research generated by these concepts, however, is relatively limited (Kissi et al., 2017) and some of the research has focused on only one of the concepts, rather than making systematic comparisons among them. For example, Zettle and Young (1987) investigated the influence of tracking on schedule performance. In particular, participants were given a computer task in which moving a marker to a point on the screen would result in earning a point, after 32 min of which an extinction phase was initiated. One group of participants were asked to report their guesses about how to respond correctly and these guesses were subsequently matched by the task contingencies, thus making their guesses always correct (tracking condition). That is, if participants reported that number of presses was the important variable, an FR-18 schedule was put in place, whereas if time passed was reported as important, a DRL-2 schedule was put in place. A second group of participants acted as controls for whom the earning of points was yoked to patterns from the tracking group. In extinction, the group in which tracking was systematically reinforced took longer to adapt to the contingencies than the yoked control group. The authors concluded that these results provided support for tracking as a functional class of rule-following.

In a separate study on pliance, Berry, Geller, Calef, and Calef (1992) investigated the extent to which drivers leaving a university car park would comply with a sign instructing them to fasten their seat belts in the presence or absence of an observer. If they followed the instruction, a message would appear thanking them for doing so (pliance). Results demonstrated that although the presence of the sign alone did increase the usage of a seat belt, the presence of an observer in conjunction with the sign increased compliance even further.

In one of the few studies that have referred to the term "augmenting" as the basis of the experimental research, Whelan and Barnes-Holmes (2004) reported a complex procedure that was designed to create a type of laboratory-induced rule, the details of which are not important at this point. The basic strategy involved using the laboratory-induced rule to establish a consequential function for a previously neutral stimulus. In effect, the procedure could be interpreted as an example of formative augmenting because the consequential functions of the stimulus emerged based on a model of instructional control, rather than through direct learning.

Although studies on pliance, tracking, and augmenting have tended to focus on only one of these classes of rule-governed behavior, there have been some attempts to explore their functional independence. As we shall see, however, the results from these studies have been far from consistent. The strategy adopted in several of these studies was to determine if sensitivity to pliance and tracking differed between participants who reported high versus low levels of depression. For example, Baruch, Kanter, Busch, Richardson, and Barnes-Holmes (2007) investigated whether the presence or absence of subclinical symptoms of depression moderated rule-following, and whether plys or tracks differentially affected this rule-following. Depressed and nondepressed individuals were given an instruction about how to perform on a subsequent matching-

to-sample (MTS) task. At first, the instructions were consistent with the contingences of the task, but after a certain number of trials an unannounced contingency reversal occurred, after which the instruction was wholly inaccurate. It is important to note that participants had been assigned to a group in which they had to read the instruction aloud to the experimenter (pliance) or privately to themselves (tracking) before exposure to the task. Participants reporting high levels of depressive symptomatology adapted more quickly to the change in schedule contingencies than did the nondepressed group. No effect was found for pliance or tracking, however.

A similar study by McAuliffe, Hughes, and Barnes-Holmes (2014) investigated the same variables: the impact of pliance and tracking on a contingency-switching MTS task, but this time in a sample of male adolescents with self-reported high versus low depressive symptomatology. As in Baruch et al. (2007), the rule initially corresponded with the programmed task contingencies and then reversed, again without warning. Likewise, participants assigned to the pliance condition were asked to read their instruction aloud to the experimenter and were told that the researcher would monitor their performance. Participants in the tracking condition, however, read their instruction privately and were given no indication that their performance would be monitored. Low-depressed participants in both the pliance and tracking conditions and high-depressed participants in the tracking condition adapted readily to the contingency change. The high-depressed participants in the pliance condition, however, persisted for longer with the now ineffective rule, a result in opposition to that reported by Baruch et al.

It is interesting that other unpublished research (Gorham, 2009) failed to replicate the findings from Baruch et al. (2007) and did produce a finding similar to McAuliffe et al. (2014), but only under specific conditions (i.e., *only* when participants retained instructions on task contingencies for the *duration* of the experiment). Therefore, it appears that the terms pliance, tracking, and augmenting have not been widely used as the basis for conducting experimental research on instructional control or rule-following, and when they have, the effects have not always been consistent. Indeed, in a recent systematic review that focused on pliance, tracking, and augmenting, it was argued that the inconsistency in the findings may be due in part to the lack of clarity in operationally defining these concepts in the first place (Kissi et al., 2017; but see Kissi, Hughes, De Schryver, De Houwer, & Crombez, 2018).

Rule-Governed Behavior and Links to Clinical Phenomena

The focus on clinical issues in the context of pliance and tracking unfolded from a general argument that emerged in the development of acceptance and commitment therapy (ACT; Hayes et al., 1999). In particular, it has been widely argued that human psychopathology (hereafter referred to as human psychological suffering) can be understood in terms of excessive rule-following, which by definition undermines or reduces contact with reinforcers in the natural environment. Once again, however, the evidence for such a claim remains extremely limited, and at least one study has produced evidence that appears contradictory.

In particular, Rosenfarb, Burker, Morris, and Cush (1993) investigated the effect of shaping versus providing rules on a contingency-switching schedule of reinforcement in individuals with and without self-reported depression. Participants responded on a

multiple DRL-1/FR-1 schedule and were either provided with a rule that accurately described responding or were provided with no such rule. After 10 DRL/FR blocks, the task contingencies reversed unbeknownst to the participants (i.e., the DRL and FR components switched). Depressed individuals who were given the rule adapted more quickly to the change in schedule contingencies than the nondepressed participants. No differences were found between depressed and nondepressed contingency-shaped groups.

On balance, a recent study with in-patients formerly presenting with delusional ideation appeared to show more persistent rule-following than nonpatient controls (Monestes, Villatte, Stewart, & Loas, 2014). In the first phase of the experiment, a left hand-side button corresponded with an FR8 component of a multiple (FR8/FI18) schedule, whereas a right hand-side button corresponded with the FI18 component. In the second phase, the contingencies were reversed; again participants were not informed about this reversal. Before responding on the schedule, participants were divided into one of three instruction groups: instructions; no instructions; or self-instructions. For the instructions group, participants were provided with instructions before the task that accurately described how to respond in Phase 1. The no-instructions group received no instructions for accurate responding. The self-instructions group received no instructions before Phase 1, but were asked to write down the best way to respond after this phase (before the onset of the reversed schedule in Phase 2). In the presence of instructions (directly given or self-produced), patients formerly presenting with delusions took longer to adapt to the reversed contingencies than did the controls. That is, when the patients received or generated instructions, they were less likely to adapt their behavior in line with the contingency reversal than were the nonclinical controls. The authors thus concluded that the results provided evidence for excessive rule-following (contingency insensitivity) in the behavior of individuals formerly presenting with delusions.

Overall, it appears that the literature on the role of rule-governed behavior in human psychological suffering is scarce and conflicted. Some research suggests that excessive rule-governance may be associated with self-reported depression and persecutory delusions (McAuliffe et al., 2014; Monestes et al., 2014). In contrast, others have reported that individuals with self-reported depression are in fact *more* sensitive to changes in environmental contingencies, or their behavior is less rule-governed (Baruch et al., 2007; Rosenfarb et al., 1993).

Summary and Conclusions

The foregoing brief history of the literature on instructional control and rules, which began with Skinner's 1966 paper, suggests that his insight led to a wealth of studies that addressed a range of issues directly relevant to human behavior. The concept of rule-governed behavior could thus be seen as being of considerable benefit, at least in the early years. That is, many interesting effects emerged from this research that helped to explain why human behavior so often differed from that of nonhuman behavior on schedules of reinforcement. On balance, when researchers began to examine rule-governed behavior for its own sake (e.g., differences in performances based on shaping contingency descriptions versus performance descriptions; shaping behavior versus instructing it), a clear rationale for doing so became less obvious. That is, at first the

research was focused on explaining why human and nonhuman behavior differed on schedules of reinforcement, but when rule-governed behavior per se became the primary focus, a core question or clearly defined research agenda appeared to be lacking (i.e., why human and nonhuman behavior differed on schedules of reinforcement). In addition, attempts to identify clear functional categories of different types of rules have generally produced inconsistent results.

Standing back now with the benefit of hindsight, the recognition that a rule (or instruction) is a commonsense concept, and lacked a clear functional-analytic basis, was a serious problem, at least for behavior analysis. Or more precisely, although researchers could, and did, attempt to identify functionally distinct classes of rule-governed behavior (e.g., pliance and tracking), a clear functional definition of what it means for a rule to specify a contingency was absent. When Sidman and his colleagues' work on equivalence relations resolved the problem of specification (i.e., exactly how rules specify contingencies; see below), this could be seen as highlighting the original lack of functional-analytic precision in the concept of rules itself. Indeed, it is perhaps for this reason that the research on rule-governed behavior appeared to be overshadowed by research on stimulus equivalence (discussed below). The remaining half of the article will explore the relationship between rule-governed behavior and derived stimulus relations, and the implications of this relationship for moving forward in this domain.

The Link between Rules and Derived Relational Responding

Another key area within the behavior-analytic literature in which human behavior has often been distinguished from that of nonhumans is derived stimulus relations. The concept of derived stimulus relations was first formalized with the seminal work of Murray Sidman (1971) and his colleagues (see Sidman, 1994, for a book-length treatment). At the time, Sidman was attempting to develop procedures for teaching basic reading skills to individuals with learning disabilities. A crucial and unexpected finding that emerged from this research was that after teaching a small number of simple relations, a number of untaught relations emerged. For example, if two abstract stimuli were matched to a third stimulus (e.g., A–B and A–C), previously unmatched and unreinforced responses often emerged (B–C and C–B). In more concrete terms, imagine that a child was presented with a picture of a lion and was taught to pick the written word "lion" and the written word "roar." In due course, the child may spontaneously match the word "lion" with the word "roar" and the word "roar" with the word "lion." When such a pattern of responses emerged, the stimuli involved were said to form an equivalence class or relation. One of the key driving factors behind the study of equivalence relations was the ease with which it was demonstrated with human participants but appeared to be largely absent (or at best extremely weak) in nonhumans. In this respect, therefore, there were now two key ways in which human and nonhuman behavior differed (i.e., rules and derived relations), and indeed, as noted above, Sidman drew upon the concept of equivalence relations to solve the issue of specification in instructional control. That is, Sidman argued that when a word participates in an equivalence relation with an object or event, this provides a functional-analytic definition of what it means for a word to specify or to refer to that object or event (see Sidman, 1994, for an extended discussion).

The extension and elaboration of the study of equivalence relations into what is now known as relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001) produced laboratory-based models of rule-governed behavior, in which Sidman's equivalence-based definition of specification was formally tested. The critical point here is that equivalence relations were treated in RFT as but one class of generalized operant behavior. According to the theory, there are numerous classes of relational operants, including difference, comparison, opposition, and temporal relations (see Hughes & Barnes-Holmes, 2016, for an extensive review). A rule or instruction, from the perspective of RFT, involves a relational network composed largely of equivalence relations among the words in the rule, and the events to which they refer, and the sequencing of the words in accordance with specific temporal relations. Thus, the simple instruction "When the light turns green, then go" involves equivalence relations among the words "light," "green," and "go" and an actual light, color, and action, with the words "when" and "then" functioning as cues for temporal relations among these events (i.e., green light *before* go).[2] According to RFT, rules or instructions involve relational networks *and* transformations of functions that provides the rule with its behavior-controlling properties. In this example, when the light turns green, it may evoke a "going" response in the absence of a direct history of reinforcement. Thus, a rule as defined by RFT extends beyond discriminative control as traditionally defined (i.e., as a stimulus with a direct reinforcement history). Indeed, a similar argument was offered by other behavior analysts in defining contingency specifying stimuli as function-altering (Blakely & Schlinger, 1987; Schlinger & Blakely, 1987). Unlike RFT, however, the definition did not include a behavioral process by which contingency-specifying stimuli acquired their function-altering effects (see Barnes-Holmes et al., 2001).

The foregoing RFT conceptual analysis of rules was first tested empirically in a study reported by O'Hora, Barnes-Holmes, Roche, and Smeets (2004) and more recently by O'Hora, Barnes-Holmes, and Stewart (2014). In the former study, participants first learned to respond to abstract stimuli as "similar," "different," "before," and "after" contextual cues. The cues were then presented with novel stimuli (nonsense syllables and colored squares) to form the functional equivalent of simple instructions, or more precisely relational networks that were shown to control specific response sequences (e.g., pressing four colored keys in a particular order: blue before red before green before yellow). O'Hora et al. (2014) replicated and extended the basic effect by showing that the derived sequence responding reported in the earlier study could itself be brought under contextual control. The critical point here is that the derived sequence responding occurred in the absence of a direct history of reinforcement, and thus provided a potential model of one of the defining features of rule-governed behavior (i.e., it allows for problem solving in the absence of direct contingency control). Overall, the research served to highlight that the simple concept of rule-governed behavior conceals what appears to be highly complex relational phenomena. The

[2] The details of RFT and the debate over its relationship with other accounts of equivalence and derived stimulus relations (e.g., naming theory) will not be discussed here, because this material would extend well beyond the scope of this article (but see the recent special issue of *Perspectives on Behavior Science*: Critchfield, Barnes-Holmes, & Dougher, 2018, for relevant articles).

remaining sections of this article will attempt to articulate the nature of this complexity and some of its implications for future research on rule-governed behavior.

The Dynamics of Derived Relational Responding as Seen through the Lens of a Hyperdimensional Multilevel Framework

Before we consider the behavioral complexities that may remain hidden when researchers rely too heavily on the concept of rules or instructions, it seems important to examine a relatively new framework that has been offered for conceptualizing the complexities involved in derived relational responding itself (see Barnes-Holmes, Barnes-Holmes, Luciano, & McEnteggart, 2017; Barnes-Holmes, Finn, Barnes-Holmes, & McEnteggart, 2018). The framework provides a hyperdimensional multilevel (HDML) "conceptual space" for analyzing the dynamics of derived relational responding and consists of five levels and four dimensions.[3] The five levels of relational development are based on conceptual and empirical analyses that have emerged from the literature on RFT (Hayes et al., 2001): (1) mutual entailment; (2) combinatorial entailment; (3) relational networks; (4) relating relations; and (5) and relating relational networks. In identifying these as levels of relational development, the HDML framework is not indicating that they are rigid or invariant "stages." Rather, lower levels (e.g., mutual entailing) are seen as containing patterns of derived relational responding that may provide an important historical context for the patterns of relational responding that occur in the levels above (e.g., combinatorial entailment). We will describe each of the levels here only briefly because they have been considered in many other sources since the publication of the seminal text on RFT (Hayes et al., 2001; see Hughes & Barnes-Holmes, 2016, for a recent detailed summary).

Mutual entailing refers to the bidirectional nature of verbal relations (e.g., if A is more than B, then B is less than A). *Relational framing,* at its simplest, involves a combination of two mutually entailed relations (e.g., if A is more than B and B is more than C, then A is more than C). *Relational networking* involves combinations of different patterns of relational framing (e.g., if A is the same as B and B is the same as C, and C is more than D, and D is more than E, then E is less than A, B, C and D). *Relating relations* involves, at its simplest, relating a mutually entailed relation to another mutually entailed relation (e.g., if A is the more than B, and in a separate relation C is more than D, then the relationship between the two relations, A > B and C > D, is the same). *Relating relational networks* is similar to relating relations, except that it applies to separate relational frames or separate complex relational networks.

The critical advance that the HDML framework provides is that the five levels of relational development are *also* divided along four dimensions: (1) coherence, (2) complexity, (3) derivation, and (4) flexibility (see Table 1). A brief description of each of the four dimensions is as follows. *Coherence* refers to the extent to which a pattern of derived relational responding coheres with previously established patterns of such responding. For instance, if an individual is informed that stimulus X is bigger than Y

[3] At first, the HDML framework was described as *multi*dimensional More recently, however, the term "hyperdimensional" has been used to reflect a more balanced emphasis on the properties of both the entailment and transformation of functions, which define derived relational responding itself (see Barnes-Holmes, 2018; Barnes-Holmes, Barnes-Holmes, & McEnteggart, 2020).

and is subsequently told that stimulus Y is smaller than X, the latter statement would likely be deemed coherent with the former. In this instance, coherence would be relatively high because the overall pattern (X > Y = Y < X) coheres so consistently with the manner in which such verbal relations have been established by the wider verbal community (e.g., there are few instances in which the statement, "if X is bigger than Y, then Y is bigger than X" would be reinforced, or not punished/corrected, by an English-speaking listener).

Complexity refers to the level of detail or density of a particular pattern of derived relational responding. For example, the mutually entailed relation of coordination may be seen as less complex than the mutually entailed relation of comparison, because the former involves only one type of relation (e.g., if X is the same as Y, then Y is the same as X), but the latter involves two types of relations (if X is bigger than Y, then Y is smaller than X).

Derivation refers to the extent to which a particular pattern of derived relational responding has previously been "practiced" or emitted. Within the HDML framework, each time a relation is derived, its derivation reduces because it acquires its own history that extends beyond the derivation that is made from the "baseline" relation. For example, imagine that an individual learns that X is bigger than Y, and thus derives that Y is smaller than X. The first time that the Y < X relation is derived, it is derived "directly" from the X > Y "baseline" relation. However, if the individual subsequently continues to respond to Y as smaller than X, that relational response gradually acquires its own history, irrespective of whether or not it is directly reinforced, rendering it less and less derived from the original baseline relation (i.e., X bigger than Y).

Flexibility refers to the extent to which a given instance of derived relational responding may be modified by current contextual variables. As a simple example, imagine a young child who is asked to respond with the wrong answer to the question, "Which is bigger, a mouse or an elephant?" The more rapidly the child responds with "mouse," the more flexible the relational responding (see O'Toole & Barnes-Holmes, 2009). Of course, flexibility is always context dependent and thus if the child had been "warned" previously not to give a wrong answer when asked to do so, it would be difficult to use the production of a correct or wrong answer as an indication of flexibility.

Table 1. A hyperdimensional multilevel (HDML) framework comprising 20 intersections between five levels and four dimensions

LEVELS	DIMENSIONS			
	Coherence	Complexity	Derivation	Flexibility
Mutually Entailing	Coh/Mut-Ent	Cpx/Mut-Ent	Der/Mut-Ent	Flx/Mut-Ent
Relational Framing	Coh/Frame	Cpx/Frame	Der/Frame	Flx/Frame
Relational Networking	Coh/Net	Cpx/Net	Der/Net	Flx/Net
Relating Relations	Coh/Rel-Rel	Cpx/Rel-Rel	Der/Rel-Rel	Flx/Rel-Rel
Relating Relational Networks	Coh/Rel-Net	Cpx/Rel-Net	Der/Rel-Net	Flx/Rel-Net

In previous publications, these units of experimental analysis have been referred to as functional-analytic abstractive relational quanta (FAARQs)

Although the HDML framework may appear to be daunting at first, it is important to appreciate that the framework simply aims to make explicit what basic researchers in RFT have been doing implicitly since the theory was first subjected to experimental analysis. That is, whenever a basic researcher in RFT conducts a study, this often involves combining at least one of the levels with one or more of the dimensions of the HDML framework. For instance, even in a simple study on equivalence relations, the researcher selects a level (e.g., mutual entailment or symmetry) and then must specify how many trials will be used to test for the entailed symmetry relations (e.g., 10), and how many trials must be "correct" to define the performance as mutual entailment (e.g., 8/10). In effect, the number of opportunities to derive the entailed relations has been specified (i.e., 10), and the number of responses that must cohere with the relations is also determined (i.e., 8). At this point, therefore, the level (mutual entailment) and two of the dimensions (derivation and coherence) of the HDML framework have been invoked. If relations other than symmetry are introduced to the study, or programmed forms of contextual control are involved, then relational complexity is also manipulated. Furthermore, if the researcher attempts to change the test performances in some manner (e.g., by altering the baseline training), then the relational flexibility in the original test performances can also be assessed.

As noted previously, researchers in the area of derived stimulus relations have been doing this type of work for decades. Thus, the HDML framework simply makes these scientific behaviors more explicit, by situating them in a framework that specifies 20 intersections between the widely recognized levels of relational development identified in RFT and the dimensions along which the levels have been or could be studied.

At this point, it seems important to emphasize that the 20 intersections identified within the HDML framework are the units of experimental analysis, whereas the levels and the dimensions per se are not. For example, although it is possible to state that mutual entailment is the bidirectional relation between two stimuli, mutual entailment can only be analyzed experimentally by specifying one or more of the dimensions. As noted previously, the tested relation must cohere in some prespecified manner with the trained relation (e.g., if X is bigger than Y, then Y will be smaller than X), and the number of derived relational responses must be specified (e.g., a participant must produce at least 8 out of 10 responses indicating that Y is indeed smaller than X in the absence of programmed reinforcement, prompting, or other feedback).

A detailed treatment of the HDML framework is beyond the scope of this article (see Barnes-Holmes, 2018; Barnes-Holmes et al., 2017; Barnes-Holmes et al., 2020, for additional details). The critical point, however, is that when the study of derived relational responding is viewed through the lens of the HDML framework, its potential in helping researchers analyze the complexities and dynamics of human language and cognition (including, most important in the current context, rule-governed behavior) may become apparent. In the next section, we will elaborate this argument by considering some recent empirical research on derived relational responding and persistence in rule-following.

Integrating the Study of Derived Relational Responding with Persistent Rule-Following: Implications Arising from the HDML Framework

Until recently, research attempting to integrate the work on derived stimulus relations and persistent rule-following was somewhat lacking. However a study by Harte,

Barnes-Holmes, Barnes-Holmes, and McEnteggart (2017) attempted to fill this gap and the interpretation of the study's results was largely aided by the HDML framework. Across two experiments, participants were given either a direct rule (direct-rule condition) or a rule that involved a novel derived relational response (derived-rule condition) that specified initially accurate responding on the subsequent MTS task. In Experiment 1, all participants first responded on 10 trials in which the rule was consistent with the task contingencies, followed by an uncued contingency reversal. Participants then responded on 50 more trials in which the rules no longer matched the contingencies. Experiment 2 partially replicated Experiment 1, but here participants were provided with 100 trials before the contingency reversal (rather than 10). Although no significant differences emerged in rule persistence between direct- and derived-rule conditions in Experiment 1, in Experiment 2, participants in the direct rule condition demonstrated significantly more persistence than did those in the derived rule condition. That is, only when participants had 100 opportunities to follow the reinforced rule (rather than only 10), did the nature of the rule affect rule persistence. Furthermore, the only correlations that emerged between rule persistence and psychological distress were in the direct-rule condition.

When viewed through the lens of the HDML framework, it could be argued that lower levels of derivation (as in the direct rule condition) produced greater levels of persistence than higher levels of derivation (as in the derived rule condition).[4] Thus, it appeared that persistence in rule-following may have varied as a function of level of derivation. This interpretation, however, does not address the fact that level of derivation appeared to have little impact on rule persistence when the opportunity to follow the reinforced rule was relatively brief (10 trials before the contingency switch in Experiment 1). When once again viewed through the lens of the HDML framework, it could be argued that coherence (between the rule and the contingencies) may interact dynamically with levels of derivation (within the rule itself) to affect persistence in rule-following. That is, when coherence between the rule and contingencies was high (100 trials in Experiment 2), level of derivation, within the rule, affected persistent rule-following; but when coherence was low, the level of derivation did not affect rule persistence. In simpler terms, participants may have been much more certain that the rule was "correct" when exposure to the contingencies was more protracted. If this interpretation is correct, it suggests that the relationship between level of derivation and rule persistence is moderated by coherence (the dynamical relationship between coherence and derivation is discussed below).

In the context of the Harte et al. (2017) study, the HDML framework allowed for a more detailed interpretation of the results than that offered by the term rule-governed behavior per se. It is critical that the HDML interpretation also suggested a future avenue of inquiry. Indeed, Harte, Barnes-Holmes, Barnes-Holmes, and McEnteggart (2018) conducted a follow-up study to test the suggestion that level of derivation, as manipulated directly within the experiment, affected rule persistence. That is, would a condition that involved *low* levels of derivation produce more persistent rule-following than a condition that involved *high* levels of derivation? The study also sought to examine the

[4] From an RFT point of view, even the direct-rule condition involved some level of derivation, given that all human verbal behavior is derived from a history of arbitrarily applicable relational responding. Thus, the direct rule condition involves a low level of derivation, whereas the derived rule condition involves a high level of derivation (see Barnes-Holmes et al., 2017, for a detailed discussion).

impact of high versus low levels of derivation across two levels of the HDML framework, that is, mutual entailment (Experiment 1) and combinatorial entailment (Experiment 2). Results showed more persistence in rule-following when derivation was low rather than high, at both levels of mutual and combinatorial entailment.

In the study by Harte et al. (2018), level of derivation was directly manipulated by providing different amounts of training. In a more recent unpublished study conducted by our research group we held derivation constant and explored the impact of coherence on persistent rule-following. Coherence was manipulated through the systematic use of feedback, based on the assumption that manipulating corrective feedback would affect level of coherence (e.g., providing feedback for "correct" derived responding would likely increase coherence). Across two experiments, all participants were first trained on the same set of baseline relations as in Harte et al. (A–B/B–C). What followed this baseline training then differed by experiment. In Experiment 1, participants were then retrained on the same baseline relations for two further blocks of trials, with one group receiving feedback on their performances and another group receiving no performance feedback. Following baseline training in Experiment 2, however, participants were directly tested on the derived A–C relations for two further blocks of trials. Once again, half of participants received feedback on their performance whereas half did not. Thus, whereas derivation was still held constant, as in Experiment 1, in Experiment 2 it was tested directly (i.e., participants were given the opportunity to derive the relations, rather than simply being retraining on the baseline relations). Thus, in principle, level of derivation was lower for Experiment 2 than Experiment 1 and was assessed in separate but related ways (i.e., retraining the baseline relations with and without feedback, versus testing the derived relations with and without feedback). Participants in both experiments then completed the contingency switching MTS task.

Manipulating coherence, through the provision versus nonprovision of feedback, affected on persistent rule-following, but only when participants were given the opportunity to derive the A–C relations (i.e., there appeared to be little if any impact for participants in Experiment 1 who were simply reexposed to the A–B and B–C relations, but there was a significant affect in Experiment 2). In effect, coherence (manipulated via feedback) appeared to have a clear impact on persistent rule-following when derivation is relatively low (i.e., is tested with or without feedback) but little if any impact when derivation is relatively high (i.e., when the baseline relations necessary for derivation were presented with or without feedback).[5] The

[5] It should be emphasized that the dimensions of coherence and derivation as conceptualized within the HDML framework may not be entirely separable. Indeed, Barnes-Holmes et al. (2017) highlighted that the boundaries among the dimensions and levels within the MDML (subsequently HDML) were "fuzzy" (p. 14). Furthermore, the main focus of the framework is to emphasize the dynamics involved in the various properties of arbitrarily applicable relational responding. Thus increases in one dimension may be seen as involving decreases in a second dimension. For example, attempting to increase coherence by providing performance feedback on a block of A–C trials would likely reduce level of derivation simply because responding on those trials itself involves deriving. Nevertheless, providing feedback versus no feedback may be one way to manipulate coherence directly while recognizing that derivation may also be affected. Although the inseparable and dynamic nature of the units specified within the HDML framework might be seen as a weakness, it is one shared with many concepts in behavior analysis, such as the relationship between the eliciting and reinforcing functions of a stimulus. Such distinctions ultimately stand or fall based on their utility within the basic science and its applications. Only time will tell if the HDML distinction between coherence and derivation, and indeed the distinctions made among the other dimensions, prove to be sufficiently useful to retain them within our scientific vocabulary.

important point here is that the subtleties in persistent rule-following, in the face of reversed reinforcement contingencies, are only now being revealed when the variables highlighted within the HDML framework are subjected to systematic analysis. The original concept of rule-governed behavior did not include specific reference to the dimensions identified within the HDML framework. Thus, one could argue that the original concept (of rule-governance) was not sophisticated enough for researchers to appreciate the subtleties involved in excessive rule-following. In our view, any account of excessive rule-following, and its involvement in human psychological suffering, will therefore be inadequate without taking on-board these subtleties.

Of course, one could argue that the empirical work outlined above could have been done without the HDML framework, and indeed, it is difficult to argue against a counterfactual. The important point, however, is the extent to which other researchers interested in complex human behavior find the framework useful, and this remains to be seen. At this point, it may be of benefit to provide a number of brief interpretations of earlier work on rules described in the first half of this article, but viewed through the lens of the HDML framework. This may further highlight the utility of the framework for identifying the complexities in performances, over and above the explanatory power of rules per se. At the very least, engaging in this exercise could help orient researchers towards areas of inquiry that may not be immediately obvious without the HDML framework. Before proceeding, we would like to emphasize that the following are simply examples of possible interpretations of earlier research on rule-governed behavior, not definitive RFT explanations. We thus offer them here simply as verbal stimuli designed to facilitate further debate, as well as conceptual and empirical analyses.

High versus low rates on FI schedule performances The earliest research on schedules of reinforcement with human participants, showing that response rates were either consistently high or low throughout an FI schedule, may be readily interpreted in terms of relational networks generated by an individual's contact with the contingencies. For example, imagine a participant who emitted only one or two responses during the first trial on an FI schedule versus a participant who emitted a large number of responses. It seems likely that two different relational networks would be generated by these separate interactions with the contingencies. In the first case, the network would likely include references to time, rather than response rate, whereas the latter network would include references to response rate rather than time. In both cases, the relational network would likely increase in coherence because responding in accordance with that network would lead to the delivery of reinforcement. Of course, the coherence of the "response-rate network" would be undermined if a participant spontaneously stopped responding for a period of time during one of the FI trials (because a reinforcer would be delivered for the first response emitted after the interval had elapsed). In this sense, the high-rate performance would not be seen as "insensitive" to the contingencies, but instead reflecting the relative coherence of a contingency-induced relational network.

Developmental differences in FI schedule performances The findings reported in the late 1970s and 1980s indicated a gradual developmental transition from the nonhuman-like performances of human infants to adult-like performances by age 5 years and above. The interpretation of this developmental trend appears relatively straightforward

in terms of the HDML framework. In particular, it seems likely that the relational abilities of infants are insufficient to allow for the generation of relatively complex relational networks as they interact with even relatively simple schedules of reinforcement. Of course, human infants would be unable even to generate complex relational networks that would function to control behavior on the schedule. As an infant matures into early childhood (between 2 and 4 years), the extent to which exposure to an FI schedule would generate the types of relational network that would control adult-like performances would increase. In particular, limited forms of relational framing would certainly be present by 2 or 3 years of age, but the types of networks that are composed of multiple interrelated relational frames, with appropriate forms of even relatively limited contextual control, would only be expected by the 5th and 6th year of development. Furthermore, the extent to which the generation of those networks would then serve to control performance on the schedule in a consistent and reliable manner would also require some experience (e.g., in technical terms, reductions in level of derivation, increases in coherence) in deriving, and responding in accordance with, relational networks in the natural environment.

The impact of interval-related stimuli on FI schedule performances Numerous studies showed that presenting various stimuli and/or tasks during FI schedules tended to induce nonhuman-like performances (i.e., scalloped patterns). Such a finding does not necessarily indicate that the stimuli and/or tasks rendered contingency-sensitive performances that were functionally identical to those observed with nonhumans. For example, presenting participants with a digital clock that counted down the seconds until the end of an FI schedule may have served to generate a relational network that controlled a behavior that topographically resembled a nonhuman performance. In effect, the relational network now included reference to the output of the clock as a source of behavioral control (e.g., "keep checking the timer before responding"). In so far as this particular network helped the participant to judge the length of the interval, the coherence of the network would increase and its derivation would decrease as the experimental session progressed.

Shaped contingency descriptions versus performance descriptions One of the key findings in this area was that shaping performance descriptions produced responding that was consistent with the descriptions rather than with the schedule. However, shaping contingency descriptions produced responding that was schedule-consistent, but inconsistent with the shaped descriptions. These findings suggest that the shaping was affecting relatively complex relational networks and these networks acquired functionally distinct controlling properties in terms of actual schedule performance. If the networks were coordinated with actual performance, less schedule sensitivity was observed than if the networks were coordinated with the schedule. When viewed through the lens of the HDML framework, therefore, the focus is on how specific relational networks are generated (i.e., through shaping) and the functional properties of those resulting networks on schedule performance. The question of schedule sensitivity per se is thus seen to be too simplistic or completely redundant.

Distinguishing between rules provided by a researcher and rules generated by participants and then shaped by the researcher The foregoing conclusion is also reflected in

another line of research. In a study reported by Hayes et al. (1986), participants were exposed to complex schedules of reinforcement (i.e., schedules that required switching from high to low rates of responding). Some participants were provided with no instructions on how to respond, whereas others received instructions on how to respond during one component of the schedule, and a third group received instructions on how to respond on both components. All participants were then exposed to a period of extinction. The key finding was that sensitivity to extinction was most pronounced when participants were provided with no rule or a rule that only described part of the schedule. This result indicates that when a relational network is generated by the participant interacting with the scheduled contingencies, rather than being fully instructed, higher levels of sensitivity (at least to extinction) are observed subsequently. This suggests that the source of a particular relational network may differ in levels of coherence, derivation, and flexibility, contingent on the original source of that network (i.e., whether it is fully or partly instructed or being generated through interactions with the scheduled contingencies). Once again, focusing on the variables highlighted in the HDML framework, as important properties of relational framing itself, appears to provide a more sophisticated understanding of how instructional control actually functions, at least in laboratory settings.

Distinguishing between different types of rule-governed behavior (pliance, tracking, and augmenting) As noted earlier, research on distinguishing different types of rules failed to yield reasonably consistent findings (see Kissi et al. 2017, for a recent review). One of the key problems, as highlighted by the authors of the recent review, was that the distinctions made among pliance, tracking, and augmenting lacked precision, at least at a functional-analytic level. However, when rules are interpreted as relational networks, that may vary along multiple dimensions (coherence, derivation, complexity, and flexibility), a more desirable level of functional-analytic precision may be achieved, one that will serve to generate more consistent findings across studies than we have managed thus far. At the very least, therefore, if the concepts of pliance, tracking, and augmenting are to be retained, the dimensions and levels specified in the HDML framework may be helpful in refining those concepts in a functional-analytic way. On balance, retaining these concepts (pliance, tracking, and augmenting) may turn out to be redundant as more sophisticated accounts of behavioral control by relational networks emerges both empirically and conceptually in the literature.

A related point, at least with regard to augmentals, is the extent to which they could be defined as *verbal* motivating operations (MOs), where "verbal" involves derived relational responding (see Poling, 2001; Poling, Lotfizadeh, & Edwards, 2017; Lotfizadeh, Edwards, & Poling, 2014, for extended discussions of the need to deal with the impact of verbal behavior on MOs). However, the disadvantage in equating the concept of an augmental with that of a verbal MO, is that the same criticisms leveled at the MO would then apply equally to augmentals (Poling et al., 2017). Indeed, as pointed out by one of the reviewers of this article, even if the two concepts are *not* equated, at least some of the criticisms of the MO continue to apply to the concept of the augmental. Furthermore, a recent attempt to update RFT has addressed the issue of motivation *without* reference to the concept of augmenting (Barnes-Holmes et al., 2020). It should be stressed that this updated version of RFT recognizes that motivation per se is an important variable, but is now incorporated into the HDML framework (see

also Gomes et al., 2019). Thus, although some researchers have previously defined augmentals as verbal MOs or verbal establishing operations/stimuli (e.g., Fagerstrøm & Arntzen, 2013; Ju & Hayes, 2008; Laraway, Snycerski, Olson, Becker, & Poling, 2014; Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O'Hora, 2002; Valdivia, Luciano, & Molina, 2006), the usefulness of doing so remains an open question given potential problems with the application of the concepts of augmentals and MOs to verbal stimuli that have motivational functions.

Summary and Conclusions

The extended quotation from Hayes et al. (1989) presented earlier in current article made the critically important point that the concept of human insensitivity (to direct contingencies) should be seen as highlighting the ubiquity and power of verbal stimuli. Almost 30 years later, that message rings far louder today than it did then. Indeed, this article is in a sense a recapitulation and an elaboration of that very message. In 1989, the study of rule-governed behavior was very much in the spotlight of behavior-analytic research and had been for some years, as evidenced by the very title of the volume from which the quotation is drawn. The study of equivalence relations was also emerging strongly in the behavior-analytic literature at that time, but the study of multiple stimulus relations was barely evident. Indeed, the first detailed treatment of multiple stimulus relations, and their potential relationship with rule-governed behavior, appeared in the same volume, and it was only 2 years later that the first empirical study appeared in the flagship journal of the discipline (Steele & Hayes, 1991). It was another 10 years before a book-length treatment of multiple stimulus relations, in the form of RFT was published (Hayes et al., 2001), which contained an account of verbal stimuli that could be used to develop a functional-analytic understanding of rule-governed behavior. An experimental analysis of how derived stimulus relations and rule-governed behavior might be combined has been slow to emerge (O'Hora et al., 2004, 2014). It is our hope that recent developments in RFT, as summarized in this article, may serve to stimulate a greater interest in defining and studying rule-governed behavior as responding in accordance with complex relational networks, along various dimensions and levels as specified within the HDML framework.

In making the foregoing arguments, we are not suggesting that the concept of rule-governed behavior be banished from behavior-analytic discourse. Rather, the concept should be seen as an important place holder or verbal stimulus that serves to orient the researcher towards a critically important domain in the experimental analysis of human behavior. Indeed, as Vaughan (1989) argued over 30 years ago, ". . . if rule-governed behavior is to be a technical term, then it is fitting for behavior analysts to argue that a functional definition is needed . . . even though others seem content with a descriptive one. . ." (p. 99).

In closing, we should emphasize that it was not our intention in this article to "destroy" the concept of rule-governed behavior, but rather to extend and refine it by bridging a gap between two areas of research within behavior analysis that have remained largely separate. In bridging that gap, it appears that increasing technical precision may be achieved by focusing on the types of variables contained in the

HDML framework that affect rule-governed behavior. In this sense, this article should be seen as an attempt to elaborate upon and extend the seminal work of two intellectual giants in the field of behavior analysis (B. F. Skinner and Murray Sidman) who gave us the concepts of rule-governed behavior and derived stimulus relations, respectively. Whether or not the increased precision we offer here allows for the concept of rule-governed behavior to be considered a full-blown technical term in behavior analysis remains to be seen.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformation. *Journal of Abnormal Psychology, 87*(1), 49–74. https://doi.org/10.1037/0021-843X.87.1.49.

Barnes-Holmes, D. (2018). The double edged sword of human language and cognition: Shall we be Olympians or fallen angels? [Blog post]. Retrieved from https://science.abainternational.org/the-double-edged-sword-of-human-language-and-cognition-shall-we-be-olympians-or-fallen-angels/rrehfeldtabainternational-org/

Barnes-Holmes, D., O'Hora, D., Roche, B., Hayes, S. C., Bissett, R. T., & Lyddy, F. (2001). Understanding and verbal regulation. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 103–117). New York, NY: Plenum.

Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., & McEnteggart, C. (2017). From IRAP and REC model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. *Journal of Contextual Behavioral Science, 6*(4), 473–483. https://doi.org/10.1016/j.jcbs.2017.08.001.

Barnes-Holmes, D., Finn, M., Barnes-Holmes, Y., & McEnteggart, C. (2018). Derived stimulus relations and their role in a behavior-analytic account of human language and cognition. *Perspectives on Behavioral Science, 41*(1), 155–173. https://doi.org/10.1007/s40614-017-0124-7.

Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2020). Updating RFT (more field than frame) and its implications for process-based therapy. *The Psychological Record.* Advanced online publication. https://doi.org/10.1007/s40732-019-00372-3.

Baruch, D. E., Kanter, J. W., Busch, A. M., Richardson, J. V., & Barnes-Holmes, D. (2007). The differential effect of instructions on dysphoric and nondysphoric persons. *The Psychological Record, 57*, 543–554. https://doi.org/10.1007/BF03395594.

Bentall, R. P., Lowe, C. F., & Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. *Journal of the Experimental Analysis of Behavior, 43*, 165–181. https://doi.org/10.1901/jeab.1985.43-165.

Berry, T. D., Geller, E. S., Calef, R. S., & Calef, R. A. (1992). Moderating effects of social assistance on verbal interventions to promote safety belt use: An analysis of weak plys. *Environment and Behavior, 24*, 653–669. https://doi.org/10.1177/0013916592245005.

Blakely, S. & Schlinger, H.D. (1987). Rules: Function-altering contingency specifying stimuli. *The Behavior Analyst, 10*(2), 183–187. https://doi.org/10.1007/BF03392428.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior, 38*, 233–248. https://doi.org/10.1901/jeab.1982.38-233.

Critchfield, T., Barnes-Holmes, D., & Dougher, M. (Eds.). (2018). Derived stimulus relations [Special Issue]. *Perspectives on Behavior Science, 41*(1).

Dixon, M., Belisle, J., Redfeldt, R., & Root, W. B. (2018). Why we are still not acting to save the world: The upward challenge of a post-Skinnerian behaviour science. *Perspectives on Behavior Science, 41*(1), 241–267. https://doi.org/10.1007/s40614-018-0162-9.

Fagerstrøm, A., & Arntzen, E. (2013). On the motivating operations at the point of online purchase setting. *The Psychological Record, 63*, 333–344. https://doi.org/10.11133/j.tpr.2013.63.2.008.

Gomes, C., Perez, W., de Almeida, J. H., Ribeiro, A., de Rose, J. C., & Barnes-Holmes, D. (2019). Assessing a derived transformation of functions using the implicit relational assessment procedure under three motivative conditions. *The Psychological Record, 69*(4), 487–497. https://doi.org/10.1007/s40732-019-00353-6.

Gorham, M. (2009). *Experimental analyses of rule-following: Methodological and clinical implications* (Unpublished doctoral dissertation). National University of Ireland Maynooth, Maynooth, Ireland.

Hantula, D. A., & Crowell, C. R. (2016). Matching and behavioral contrast in a two-option repeated investment simulation. *Managerial & Decision Economics, 37*, 294–305. https://doi.org/10.1002/mde.2717.

Hantula, D. A., Brockman, D. D., & Smith, C. L. (2008). Online shopping as foraging: The effects of increasing delays on purchasing and patch residence. *IEEE Transactions on Professional Communication, 51*, 147–154. https://doi.org/10.1109/tpc.2008.2000340.

Harte, C., Barnes-Holmes, Y., Barnes-Holmes, D., & McEnteggart, C. (2017). Persistent rule-following in the face of reversed reinforcement contingencies: The differential impact of direct versus derived rules. *Behavior Modification, 41*(6), 743–763. https://doi.org/10.1177/0145445517715871.

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2018). The impact of high versus low levels of derivation for mutually and combinatorially entailed relations on persistent rule-following. *Behavioural Processes, 157*, 36–46. https://doi.org/10.1016/j.beproc.2018.08.005.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control*. New York, NY: Plenum.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153–190). New York, NY: Plenum.

Hayes, S. C., & Wolf, M. R. (1984). Cues, consequences, and therapeutic talk: Effect of social context and coping statements on pain. *Behavior Research & Therapy, 22*, 385–392. https://doi.org/10.1016/0005-7967(84)90081-0.

Hayes, S. C., Rosenfarb, I., Wulfert, E., Munt, E. D., Korn, Z., & Zettle, R. D. (1985). Self-reinforcement effects: An artifact of social standard setting? *Journal of Applied Behavior Analysis, 18*(3), 201–214. https://doi.org/10.1901/jaba.1985.18-201.

Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, multiple schedules, and extinction: Distinguishing rule-governed from schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 46*, 137–147. https://doi.org/10.1901/jeab.1986.46-137.

Hayes, S. C., Zettle, R. D., & Rosenfarb, I. (1989). Rule-following. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 191–220). New York, NY: Plenum.

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Plenum.

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 129–178). West Sussex, UK: Wiley.

Ju, W. C., & Hayes, S. C. (2008). Verbal establishing stimuli: Testing the motivative effect of stimuli in a derived relation with consequences. *The Psychological Record, 58*, 339–363. https://doi.org/10.1007/BF03395623.

Kissi, A., Hughes, S., Mertens, G., Barnes-Holmes, D., De Houwer, J., & Crombez, G. (2017). A systematic review of pliance, tracking, and augmenting. *Behavior Modification, 41*(5), 683–707. https://doi.org/10.1177/0145445517693811.

Kissi, A., Hughes, S., De Schryver, M., De Houwer, J., & Crombez, G. (2018). Examining the moderating impact of plys and tracks on the insensitivity effect: A preliminary investigation. *The Psychological Record, 68*, 431–440. https://doi.org/10.1007/s40732-018-0286-z

Laraway, S., Snycerski, S., Olson, R., Becker, B., & Poling, A. (2014). The motivating operations concept: Current status and critical response. *The Psychological Record, 64*(3), 601–623. https://doi.org/10.1007/s40732-014-0080-5.

Laties, V. G., & Weiss, B. (1963). Effects of a concurrent task on fixed-interval responding in humans. *Journal of the Experimental Analysis of Behavior, 6*(3), 431–436. https://doi.org/10.1901/jeab.1963.6-431.

Leander, J. D., Lippman, L. G., & Meyer, M. M. (1968). Fixed interval performance as related to instructions to subjects' verbalization of the reinforcement contingency. *The Psychological Record, 18*, 469–474. https://doi.org/10.1007/BF03393795.

Lippman, L. G., & Meyer, M. M. (1967). Fixed interval performance as related to instructions to subjects' verbalization of the contingency. *Psychonomic Science, 8*, 135–136. https://doi.org/10.3758/BF03331586.

Lotfizadeh, A. D., Edwards, T., & Poling, A. (2014). Motivating operations in the *Journal of Organizational Behavior Management*: Review and discussion of relevant articles. *Journal of Organizational Behavior Management, 34*, 69–103. https://doi.org/10.1080/01608061.2014.914010.

Lowe, C. F., Harzem, P., & Bagshaw, M. (1978a). Species differences in temporal control of behavior II: Human performance. *Journal of the Experimental Analysis of Behavior, 29*(3), 351–361. https://doi.org/10.1901/jeab.1978.29-351.

Lowe, C. F., Harzem, P., & Hughes, S. (1978b). Determinants of operant behavior in humans: Some differences in animals. *Quarterly Journal of Experimental Psychology, 30*(2), 373–386. https://doi.org/10.1080/14640747808400684.

Lowe, C. F., Beasty, A., & Bentall, R. P. (1983). The role of verbal behavior in human learning: Infant performance on fixed interval schedules. *Journal of the Experimental Analysis of Behavior, 39*, 157–164. https://doi.org/10.1901/jeab.1983.39-157.

Matthews, B. A., Catania, A. C., & Shimoff, E. (1985). Effects of uninstructed verbal behavior on non-verbal responding: Contingency descriptions versus performance descriptions. *Journal of the Experimental Analysis of Behavior, 43*, 155–164. https://doi.org/10.1901/jeab.1985.43-155.

McAuliffe, D., Hughes, S., & Barnes-Holmes, D. (2014). The dark-side of rule governed behavior: An experimental analysis of problematic rule-following in an adolescent population with depressive symptomatology. *Behavior Modification, 38*(4), 587–613. https://doi.org/10.1177/0145445514521630.

Monestes, J. L., Villatte, M., Stewart, I., & Loas, G. (2014). Rule-based insensitivity and delusion maintenance in schizophrenia. *The Psychological Record, 64*(2), 329–338. https://doi.org/10.1007/s40732-014-0029-8.

O'Hora, D., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2004). Derived relational networks and control by novel instructions: A possible model of generative verbal responding. *The Psychological Record, 54*, 437–460. https://doi.org/10.1007/BF03395484.

O'Hora, D., Barnes-Holmes, D., & Stewart, I. (2014). Antecedent and consequential control of derived instruction-following. *Journal of the Experimental Analysis of Behavior, 102*(1), 66–85. https://doi.org/10.1002/jeab.95.

O'Toole, C. & Barnes-Holmes, D. (2009). Three chronometric indices of relational responding as predictors of performance on a brief intelligence test: The importance of relational flexibility. *The Psychological Record, 59*, 119–132. https://doi.org/10.1007/BF03395652.

Poling, A. (2001). Commentary regarding Olson, Laraway, and Austin (2001). *Journal of Organizational Behavior Management, 21*, 47–56. https://doi.org/10.1300/J075v21n02_06.

Poling, A., Lotfizadeh, A., & Edwards, T. L. (2017). Predicting reinforcement: Utility of the motivating operations concept. *The Behavior Analyst, 40*(1), 49–56. https://doi.org/10.1007/s40614-017-0091-z.

Roche, B., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & O'Hora, D. (2002). Relational frame theory: A new paradigm for the analysis of social behavior. *The Behavior Analyst, 25*(1), 75–91. https://doi.org/10.1007/BF03392046.

Rosenfarb, I. S., Burker, E. J., Morris, S. A., & Cush, D. T. (1993). Effects of changing contingencies on the behavior of depressed and nondepressed individuals. *Journal of Abnormal Psychology, 102*(4), 642–646. https://doi.org/10.1037/0021-843X.102.4.642.

Schlinger, H.D. & Blakely, S. (1987). Function-altering effects of contingencyspecifying stimuli. *The Behavior Analyst, 10,* 41–45. https://doi.org/10.1007/BF03392405.

Seligman, M. E. P. (1974). Depression and learned helplessness. In R. J. Friedman & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 83–126). Washington, DC: Winston-Wiley.

Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior, 36*, 201–220. https://doi.org/10.1901/jeab.1981.36-207.

Shimoff, E., Matthews, B. A., & Catania, A. C. (1986). Human operant performance: Sensitivity and pseudosensitivity to contingencies. *Journal of the Experimental Analysis of Behavior, 46*, 149–157. https://doi.org/10.1901/jeab.1986.46-149.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, & Hearing Research, 14*, 5–13. https://doi.org/10.1044/jshr.1401.05.

Sidman, M. (1994). *Equivalence relations and behaviour: A research story*. Boston, MA: Authors Cooperative.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Keinmuntz (Ed.), *Problem-solving: Research, method, and therapy* (pp. 225–257). New York, NY: Wiley.

Steele, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56*, 519–555. https://doi.org/10.1901/jeab.1991.56-519.

Torneke, N., Luciano, C., & Valdivia-Salas, S. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology & Psychological Therapy, 8*(2), 141–156.

Valdivia, S., Luciano, C., & Molina, F. J. (2006). Verbal regulation of motivational states. *The Psychological Record, 56*, 577–595. https://doi.org/10.1007/BF03396035.

Vaughan, M. (1989). Rule-governed behavior in behavior analysis: A theoretical and experimental history. In S.C. Hayes (Ed.), Rule-governed behavior: Cognition, contingencies and instructional control (pp. 97-118). New York, NY: Plenum.

Weiner, H. (1964). Conditioning history and human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 7*, 383–385. https://doi.org/10.1901/jeab.1964.7-383.

Weiner, H. (1969). Controlling human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 12*, 349–373. https://doi.org/10.1901/jeab.1969.12-349.

Whelan, R., & Barnes-Holmes, D. (2004). Empirical models of formative augmenting in accordance with the relations of same, opposite, more-than, and less-than. *International Journal of Psychology & Psychological Therapy, 4*, 285–302.

Zettle, R. D., & Hayes, S. C. (1982). Rule-governed behavior: A potential theoretical framework for cognitive-behavior therapy. In P. C. Kendall (Ed.), *Advances in cognitive-behavioral research and therapy Vol. 1* (pp. 73–118). New York, NY: Academic Press.

Zettle, R. D., & Young, M. J. (1987). Rule-following and human operant responding: Conceptual and methodological considerations. *Analysis of Verbal Behavior, 5*(1), 33–39. https://doi.org/10.1007/BF03392818.