Perspect Behav Sci
DOI 10.1007/s40614-017-0124-7

ORIGINAL RESEARCH

Derived Stimulus Relations and Their Role in a Behavior-Analytic Account of Human Language and Cognition

Dermot Barnes-Holmes[1] · **Martin Finn**[1] · **Ciara McEnteggart**[1] · **Yvonne Barnes-Holmes**[1]

Abstract This article describes how the study of derived stimulus relations has provided the basis for a behavior–analytic approach to the study of human language and cognition in purely functional–analytic terms, with a focus on basic rather than applied research. The article begins with a brief history of the early behavior–analytic approach to human language and cognition, focusing on Skinner's (1957) text *Verbal Behavior*, his subsequent introduction of the concept of instructional control (Skinner, 1966), and Sidman's (1994) seminal research on stimulus equivalence relations. The article then considers how the concept of derived stimulus relations, as conceptualized within relational frame theory (Hayes et al., 2001), allowed researchers to refine and extend the functional approach to language and cognition in multiple ways. Finally, the article considers some recent conceptual and empirical developments that highlight how the concept of derived stimulus relations continues to play a key role in the behavior–analytic study of human language and cognition, particularly implicit cognition. In general, the article aims to provide a particular perspective on how the study of derived stimulus relations has facilitated and enhanced the behavior analysis of human language and cognition, particularly over the past 25–30 years.

Keywords Derived stimulus relations · Relational frame theory · Human · Language · Cognition

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

✉ Dermot Barnes-Holmes
Dermot.Barnes-Holmes@ugent.be

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

Published online: 14 November 2017

Arguing that behavior analysis has developed an approach to human language and cognition could be seen as paradoxical. These two concepts are typically seen as inherently mentalistic; thus, behavior analysis, which is explicitly monistic, surely cannot provide a scientific account of such phenomena. We certainly agree with this view when language and cognition are treated as mentalistic concepts. For example, if a given instance of human language is seen as reflecting or capturing a mental event (i.e., a cognition) that represents an individual's understanding of the external world, behavior analysis would have little, if anything, to say about such a statement (see De Houwer, 2017, for a recent detailed discussion of the relationship between behavior analysis and cognitive psychology, with a focus on stimulus relations research). If, however, language and cognition are treated, at least initially, as ill-defined domains or areas of human behavior that may be subjected to monistic, functional analyses, we can see no reason to object to the use of the terms per se (see Hayes, 1984). The question that remains, of course, is how best to subject the behaviors we typically associate with human language and cognition to systematic functional analysis. An obvious and straightforward answer to this question was not immediately forthcoming in the history of the discipline, but we will argue that it was the study of stimulus relations, and derived stimulus relations in particular, that provided an important advance in this regard. In other words, when derived stimulus relations were defined as core functional–analytic units of human language and cognition, a whole range of both basic and applied research possibilities became apparent. In making this argument, we are not suggesting that the concept of derived stimulus relations provides the only way in which behavior analysis can or should study and explain human language and cognition. Rather, we are simply suggesting that the concept has proven to be a particularly useful one in this regard.

The purpose of this article is to highlight how the study of derived stimulus relations provided the basis for a functional analysis of human language and cognition. The article will begin with a brief history of the early behavior–analytic approach to these domains (see Hayes, 1989, for a book-length treatment). It will then consider how the concept of derived stimulus relations allowed researchers to refine and extend the approach in multiple ways (see Hayes, Barnes-Holmes, & Roche, 2001, and Sidman, 1994, for detailed narratives). Finally, we will consider some very recent conceptual and empirical developments that highlight how the concept of stimulus relations continues to play a key role in the behavior–analytic study of human language and cognition, particularly so-called implicit cognition.

A Brief Historical Review

Before continuing, we should emphasize that what we present is not meant to be a comprehensive and completely balanced historical review of the history of basic behavior–analytic research on human language and cognition. Instead, what we offer is a particular perspective on how the study of derived stimulus relations has facilitated and enhanced the behavioral study of human language and cognition, particularly over the past 25–30 years.

Skinner's *Verbal Behavior*, Instructional Control, and Schedule Insensitivity: the Early Beginnings of a Functional Approach to Human Language and Cognition

Not surprisingly, the initial development of a behavior–analytic account of human language and cognition was provided by the discipline's main progenitor, B. F. Skinner. Specifically, Skinner (1957) published a book-length treatment of human language, *Verbal Behavior*, and subsequently published an account of human problem solving that appealed to the concept of instructional control or rule-governed behavior (Skinner, 1966). In *Verbal Behavior*, Skinner proposed a range of verbal operants, such as mands, tacts, and intraverbals. For example, the concept of the tact defined an object, such as an apple, as discriminative for emitting the tact "apple" when doing so had previously been reinforced by a listener or listeners in the verbal community. Instructions were defined as stimuli that specified reinforcement contingencies and could thus be used to solve problems without a listener having to contact the contingencies directly. These original behavior–analytic approaches to human language and cognition did not draw upon the concept of derived stimulus relations. For example, an apple and the tact "apple" were not seen as participating in a derived stimulus relation; rather, the object was defined as a discriminative stimulus for the tact and no more. Of course, the absence of derived relations in Skinner's proposed verbal operants is hardly surprising given that the seminal work on such relations did not emerge until the 1970s. Indeed, it is worth noting that when *Verbal Behavior* was being composed (mostly during the 1940s), research on conditional discriminations, which eventually spawned stimulus relations work, had not even begun. It is to Skinner's credit, therefore, that he did in fact provide some examples of derived stimulus relations in *Verbal Behavior* (Moore, 2009), but at the time he lacked the empirical guidance to incorporate such relations directly into many of the verbal operants he proposed in the book (but see Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000, for one attempt to synthesize Skinner's verbal operants with derived stimulus relations). It should also be acknowledged that research arising from *Verbal Behavior* has had, and continues to have, a very significant impact on the assessment and treatment of language deficits in developmentally delayed populations (e.g., McKeel, Rowsey, Belisle, Dixon, & Szekely, 2015).

Although Skinner provided the initial conceptual foundation for a behavior–analytic approach to human language and cognition, developments in the experimental wing of the discipline were also highly relevant. Specifically, differences in the behavior of humans and nonhumans when they were exposed to schedules of reinforcement (Bentall, Lowe, & Beasty, 1985; Lowe, Beasty, & Bentall, 1983; Weiner, 1969) suggested that the development of human language created important species differences (Lowe, 1979). The basic argument was that some form of precurrent behavior, typically conceptualized as verbal, affected responding on a reinforcement schedule (e.g., Catania, Shimoff, & Matthews, 1989) and rendered human behavior less sensitive to the contingencies. Often, the so-called insensitivity effect observed with human schedule performance was attributed to the impact of *instructions* or *rules* that were generated by human participants as they interacted with the scheduled contingencies (e.g., Vaughan, 1989). Insofar as nonhumans did not possess the capacity for generating such rules, their behavior was seen as being directly controlled by—or entirely sensitive to—schedules of reinforcement.

As noted earlier, Skinner (1966) himself introduced the concept of instructions or rules to the behavior–analytic tradition in a seminal paper on human problem solving, and thus the focus on rule-governed behavior in the context of human schedule performance was not considered particularly problematic. On balance, the introduction of the concept of rule-governed behavior as a means of explaining human insensitivity to reinforcement schedules served to highlight a problem with a core assumption in behavior–analytic thinking. Specifically, the experimental and conceptual analyses of behavior that had been wrought from decades of work with nonhumans did not seem to apply, at least in whole cloth, to the behavior of verbally sophisticated humans (Dymond, Roche, & Barnes-Holmes, 2003). One obvious option was simply to abandon the "continuity" (between nonhuman and human learning) assumption and build a basic science of human behavior that continued to draw on nonhuman principles of behavior but was not constrained by them. To a certain extent, this is the approach that some researchers began to adopt by studying, for example, the impact of different types of rules on human schedule performance (e.g., Catania et al., 1989). In other words, the focus shifted from asking why human schedule performances often differed from those of nonhumans to analyzing the impact of rules and instructions per se. Once again, it seems important to note that the concept of derived stimulus relations did not play a key or pivotal role in the research on human schedule performance and the effects of instructions or rules. Indeed, for many years, if not decades, the study of rule-governed behavior continued with little or no connection to the study of derived stimulus relations. Eventually, the two would cross paths, but that took almost 20 years.

Stimulus Equivalence and Relational Frame Theory: Derived Stimulus Relations as Core Units of Human Language and Cognition

The concept of and research on derived stimulus relations can be traced back to the seminal work of Murray Sidman on what he called "stimulus equivalence" (e.g., Sidman, 1971, 1994; Sidman & Tailby, 1982). Interestingly, the earliest work in the area started with Sidman and colleagues' efforts to develop methods for teaching basic reading skills. Thus, from the very beginning, the study of derived stimulus relations was focused on human language. The basic stimulus equivalence effect was defined as the emergence of unreinforced or untrained matching responses based on a small set of trained responses. For example, when a person was trained to match two abstract stimuli to a third (e.g., *Zid-Paf* and *Zid-Vek*), untrained matching responses frequently appeared in the absence of additional learning (e.g., *Paf-Vek* and *Vek-Paf*). When such a pattern of unreinforced responses occurred, the stimuli were said to form an equivalence class or relation. Importantly, this behavioral effect appeared to provide the basis for a functional–analytic definition of symbolic meaning or semantic reference (see Sidman, 1994). In other words, a written or spoken word could only be defined as a symbol for an object or event if it participated in an equivalence class with that other stimulus. Within behavior analysis, this could be seen as a major breakthrough in terms of defining symbolic meaning in functional terms, in part because it paved the way for a broad research program that incorporated many areas of human language and cognition that before then had remained largely untouched in the basic science of behavior analysis.

The key extension to the work on stimulus equivalence arrived in the form of relational frame theory (RFT; Hayes, 1991; Hayes et al., 2001). Specifically, the theory

argued that stimulus equivalence may be considered a generalized relational operant and that many different classes of such operants were possible and indeed common in natural human language. According to this view, exposure to an extended history of relevant reinforced exemplars served to establish particular patterns of generalized relational (operant) response units, defined as relational frames (Barnes-Holmes & Barnes-Holmes, 2000). For example, a young child would likely be exposed to direct contingencies of reinforcement by the verbal community for pointing to the family dog upon hearing the word *dog* or the specific dog's name (e.g., *Rover*) and to emit other appropriate naming responses, such as saying *Rover* or *dog* when the family pet was observed or saying *Rover* when asked *What is the dog's name?* Across many such exemplars involving other stimuli and contexts, eventually the operant class of coordinating stimuli is established in this way, such that direct reinforcement for all of the individual components of naming is no longer required when novel stimuli are encountered. Imagine, for example, that the child was shown a picture of an aardvark and the written word and was told its name. Subsequently, the child may say *That's an aardvark* when presented with a relevant picture or the word without any prompting or direct reinforcement for doing so. In other words, once the generalized relational response of coordinating pictorial stimuli, spoken stimuli, and written words is established, directly reinforcing a subset of the relating behaviors "spontaneously" generates the complete set. Critically, when this pattern of relational responding has been established, the generalized relational response may then be applied to any stimuli given appropriate contextual cues.[1]

Contextual cues were thus seen as functioning as discriminative for particular patterns of relational responding. The cues acquired their functions through the types of histories described previously. Thus, for example, the phrase *that is a*, as in *That is a dog*, would be established across exemplars as a contextual cue for the complete pattern of relational responding (e.g., coordinating the word *dog* with actual dogs). Similarly, phrases such as *That is not a*, *That is bigger than*, or *That is faster than* would be established across exemplars as cues for other patterns (or frames) of relational responding. Once the relational functions of such contextual cues are established in the behavioral repertoire of a young child, the number of stimuli that may enter into such relational response classes becomes almost infinite.

Contextual cues were also seen as critical in controlling the behavioral functions of the stimuli that are evoked in any instance in which stimuli are related. For example, the word *dog* could evoke different responses for a child if she were asked *What does your dog look like?* versus *What does your dog smell like?* In RFT, therefore, two broad classes of contextual cues are involved in any instance of relational framing—one class controls the type of relation (e.g., equivalence), and the other cue controls the specific behavioral functions of the stimuli that are evoked during the act of relating; these two classes of contextual cues are denoted *Crel* and *Cfunc*, respectively.

The core analytic unit of the relational frame was defined as possessing three properties: mutual entailment (if A is related to B, then B is also related to A); combinatorial entailment (if A is related B and B is related to C, then A is related to

[1] According to RFT, it is the exemplar training that is critical in establishing derived relational responding, not naming per se (see Luciano, Becerra, & Valverde, 2007); naming is seen as just one way in which multiple-exemplar training may occur in the natural verbal environment.

C and C is related to A); and the transformation of functions (the functions of the related stimuli are changed or transformed based upon the types of relations into which those stimuli enter). The third property—the transformation of functions—marked a substantive and important extension to the concept of equivalence relations. Specifically, it ensured that the concept of the relational frame would always refer to some change or modification in the behavioral functions of the framed stimuli that extended beyond their relational functions per se. For example, if A was *less than* B and B was *less than* C in a particular instance of relational framing and a reinforcing function was attached to A, then C may acquire a greater reinforcing function than A or B. The concept of a relational frame was thus designed to capture how human language and cognition changes our reactions to the "real" world around us rather than simply providing, for example, an analysis of logical or abstract human reasoning.

According to RFT, many of the functions of stimuli that we encounter in the natural environment may appear to be relatively basic or simple but have acquired those properties due, at least in part, to a history of relational framing. Even a simple tendency to orient more strongly toward one stimulus than another in your visual field may be based on relational framing. Identifying the name of your hometown or city from a random list of place names may occur more quickly or strongly because it coordinates with other stimuli that control strong orienting functions (e.g., the many highly familiar stimuli that constitute your hometown). Such functions may be defined as Cfunc properties because they are examples of specific stimulus functions (i.e., orienting) that are acquired based on (but are separate from) the entailed relations among the relevant stimuli.

Following the initial exposition of RFT, the 1990s and 2000s saw a period of demonstration research that was designed to test its basic assumptions and core ideas. Some of this early research showed that relational framing, or arbitrarily applicable relational responding (AARRing), as a process can be shown to occur in several distinct patterns. These patterns, referred to as relational frames (e.g., coordination, opposition, distinction, comparison, spatial frames, temporal frames, deictic relations, and hierarchical relations), were demonstrated across numerous experimental studies (see Hughes & Barnes-Holmes, 2016a, for a recent review), and some of the research also reported reliable demonstrations of the property of the transformation of functions (e.g., Dougher, Hamilton, Fink, & Harrington, 2007; Dymond & Barnes, 1995; Roche & Barnes, 1997). In addition, research showed that relational framing could be observed using a variety of procedures (e.g., Barnes, Smeets, & Leader, 1996), indicating that the phenomenon was not tied to particular experimental preparations or modes of instruction, provided that the key functional elements were present. Empirical evidence also emerged to support the argument that exposure to multiple exemplars during early language development is required to establish specific relational frames (e.g., Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004; Lipkens, Hayes, & Hayes, 1993; Luciano, Becerra, & Valverde, 2007). As such, the argument that relational frames may be thought of as overarching or generalized relational operants (i.e., established by appropriate multiple exemplars) gained considerable traction (see Barnes-Holmes & Barnes-Holmes, 2000; Healy, Barnes-Holmes, & Smeets, 2000).

The seminal text on RFT also used the basic operant unit of the relational frame to provide functional–analytic accounts of specific domains of human language and cognition, and rule-governed behavior was one of these domains (Barnes-Holmes,

O'Hora, Roche, Hayes, Bissett, & Lyddy, 2001). According to RFT, a rule or instruction may be considered a network of relational frames typically involving coordination and temporal relations with contextual cues that transform specific behavioral functions. Take this simple instruction, for example: *If the light is green, then go*. This rule involves frames of coordination between the words *light*, *green*, and *go* and the actual events to which they refer. In addition, the words *if* and *then* serve as contextual cues for establishing a temporal relation between the green light and the act of going (i.e., first green light, then go). The relational network thus transforms the functions of the green light itself, such that it now controls the act of "going" whenever an individual who was presented with the rule observes the green light being switched on.

Additional conceptual developments generated experimental and applied analyses of verbal rules or instructions in terms of complex relational networks composed of multiple relational frames (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004), analogical and metaphorical reasoning in terms of relating relational frames (e.g., Barnes, Hegarty, & Smeets, 1997), and problem solving in terms of increasingly complex forms of contextual control over relational framing itself (Hayes, Gifford, Townsend, & Barnes-Holmes, 2001). To illustrate, consider the example of an analogy: *Pear is to peach as cat is to dog*. In this example, there are two relations coordinated through class membership (controlled by the cue *is to*) and a coordination relation that links the two coordination relations (controlled by the cue *as*). From an RFT point of view, analogical reasoning thus involves the same behavioral process involved in relational framing more generally (i.e., AARRing), but applied to framing itself (see Stewart & Barnes-Holmes, 2004).

RFT research has also focused, both conceptually and empirically, on the role of human language in the development of self. For RFT, a basic "verbal" self involves three deictic relations[2]: (a) the interpersonal relations I–YOU, (b) the spatial relations HERE–THERE, and (c) the temporal relations NOW–THEN (Barnes-Holmes, 2001). The core postulate here is that as children learn to respond in accordance with these relations, they are in essence learning to relate the self to others in the context of particular times and places. Imagine a very young child who is asked "What did you have for lunch today?" while he or she is eating an evening meal with his or her family. If the child responded simply by referring to what a sibling is currently having for dinner, he or she may well be corrected with "No, that's what your brother is eating now, but what did you eat earlier today?" In effect, this kind of ongoing refinement of the three deictic relations allows the child to respond appropriately to questions about his or her own behavior in relation to others, as it occurs at specific times and in specific places (e.g., McHugh, Barnes-Holmes, & Barnes-Holmes, 2004).

Some Recent Advances

At this point, we could continue to provide many examples of ways in which RFT has been used to provide functional accounts and approaches to various domains in

[2] The term *deictic* is used here to refer to verbal relations that specify an individual as located in a particular space (e.g., "here" rather than "there") and at a particular time (e.g., "now" rather than "then").

psychology, including intelligence, implicit cognition, prejudice, and so on (see Hughes & Barnes-Holmes, 2016b). At a more general level, however, it may be useful to consider a recently proposed framework that highlights the potential of RFT to take a class of behavior called AARRing and construct increasingly complex analyses of human language and cognition in purely functional terms. Specifically, researchers have recently offered what they describe as a multidimensional, multilevel (MDML) framework for analyzing AARRing (Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, 2016). According to this framework, AARRing may be conceptualized as developing in a broad sense from (a) mutual entailment to (b) simple networks involved in frames to (c) more complex networks involved in rules and instructions to (d) the relating of relations involved in analogical reasoning, and finally to (e) relating relational networks, which is typically involved in complex problem solving. In identifying these as levels of relational development, the MDML framework is not indicating that they are rigid or invariant "stages." Rather, lower levels (e.g., mutual entailing) are seen as containing patterns of AARRing that may provide an important historical context for the patterns of AARRing that occur in the levels above (e.g., relational framing). In general, the different levels are based on a combination of well-established assumptions within RFT and, where possible, empirical evidence. The framework also conceptualizes each of these levels as having multiple dimensions: coherence, complexity, derivation, and flexibility. Each of the levels is seen as intersecting with each of the dimensions, yielding a framework that consists of 20 units of analysis for conceptualizing and studying the dynamics of AARRing in the laboratory and in natural settings (see Table 1 for an overview of the structure of the MDML framework).

A brief description of each of the four dimensions is as follows. Coherence refers to the extent to which AARRing is generally predictable based on prior histories of reinforcement. For example, the statement *A mouse is larger than an elephant* would typically be seen as lacking coherence with the relational networks that operate in the wider verbal community. Note, however, that such a statement may be seen as coherent in certain contexts (e.g., when playing a game of "everything is opposite"). Complexity refers to the level of detail or density of a particular pattern of AARRing. As a simple 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 A is the same as B, then B is the same as A), but the latter

Table 1 A Multidimensional, Multilevel (MDML) Framework Consisting of 20 Intersections Between the Dimensions and Levels of Arbitrarily Applicable Relational Responding

Level	Dimension			
	Coherence	Complexity	Derivation	Flexibility
Mutually entailing	Analytic Unit 1	Analytic Unit 2	—	—
Relational framing	—	—	—	—
Relational networking	—	—	—	—
Relating relations	—	—	—	—
Relating relational networks	—	—	—	Analytic Unit 20

involves two types of relations (if A is bigger than B, then B is smaller than A).[3] Derivation refers to how well practiced a particular instance of AARRing has become. Specifically, when a pattern of AARRing is derived for the first time, it is, by definition, highly derived (i.e., novel or emergent), and thus derivation reduces as that pattern becomes more practiced. Flexibility refers to the extent to which a given instance of AARRing may be modified by current contextual variables. 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 easily this is achieved, the more flexible the AARRing.

Although the MDML framework may appear to be quite 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 MDML framework. 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 and two of the dimensions of the MDML 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 is also assessed. As noted previously, RFT researchers—and to some extent stimulus equivalence researchers before them—have been doing this type of work for decades. Thus, the MDML 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 well-established 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 MDML framework specify the units of *experimental* analysis, not the levels or the dimensions per se. 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 A is bigger than B, then B will be smaller than A), 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 B is indeed smaller than A in the absence of programmed reinforcement, prompting, or other feedback).

A detailed treatment of the MDML framework is beyond the scope of this article. The critical point to appreciate, however, is that RFT may be used to generate a conceptual

[3] Relational complexity itself may be defined along more than one dimension, such as the number of relata, frames, and/or contextual cues in a network. In some cases, therefore, identifying a single continuum of relational complexity may require appropriate multidimensional scaling (e.g., Borg & Groenen, 2005).

framework that begins with a very simple or basic scientific unit of analysis: the mutually entailed derived stimulus relation. From an RFT perspective, this unit is not synonymous with naming, in a casual or informal sense of that term, but it is seen to be intrinsic to it in a psychological analysis of naming as an act in context. In other words, the concept of mutual entailment suggests that learning to name likely involves, inter alia, the relational processes that define the unit of mutual entailment itself. What the MDML framework adds to this conceptual analysis is a framework for considering what appear to be the key dimensions along which mutual entailment as a behavioral process may vary (e.g., mutually entailed responding may vary in terms of coherence, complexity, derivation, and flexibility). In addition, the MDML framework emphasizes that more complex units of analysis may evolve from mutual entailment, such as the simple relational networks involved in relational frames, more complex networks involving combinations of frames, the relating of relational frames to relational frames, and ultimately the relating of entire complex relational networks to other complex relational networks. And in each case, these different levels of AARRing may vary along the four aforementioned dimensions and perhaps even others that remain to be identified.

When RFT is viewed through the lens of the MDML framework, its potential in helping researchers to analyze the complexities and dynamics of human language and cognition may become apparent. In much the same way that mutual entailment provides a purely relational approach to understanding naming as a language process, the concepts of frames, networks, relating relations, and relating relational networks provide purely relational analyses of increasingly complex human language phenomena. As outlined previously, for example, the concept of relating relations appears to be relevant to, if not synonymous with, analogical reasoning. Similarly, relating relational networks may be relevant to the telling and understanding of complex stories (Stewart, Barnes-Holmes, Hayes, & Lipkens, 2001).

We must also acknowledge that the MDML framework is a relatively new development in the RFT literature. Critically, however, the MDML framework emerged from a highly active empirical program of RFT-based research that drew heavily on the concept of stimulus relations in developing a behavior–analytic approach to so-called implicit cognition. We will briefly review this line of research because it exemplifies how the concept of stimulus relations continues to play a key role in the experimental and conceptual analysis of human language and cognition.

The Dynamics of Relational Framing: the Need for New Procedures

Much of the early research in RFT consisted of "demonstration of principle" studies that were designed to test the theory's basic assumptions and core ideas. One of the defining features of this demonstration research was a dichotomous approach to AARRing itself. In other words, basic laboratory studies in RFT often focused on showing that particular patterns of AARRing were either present or absent. Thus, for example, participants were required to produce perhaps 18 out of 20 correct responses on a test for equivalence or coordination responding to demonstrate that the relational frame had emerged. In this sense, the frame was either present or absent in the participant's behavioral repertoire. A critical feature of the concept of operant behavior generally, however, is that it may vary in relative strength. Thus, for example, the simple operant of lever pressing for food pellets in rats may be at relatively high or low strength. One way in which researchers

have typically assessed such strength is by measuring how long it takes for the operant to extinguish when the reinforcement contingency (between lever pressing and food pellets) is terminated. In effect, the longer the extinction process takes, the stronger the operant response class may be deemed to be.

Basic RFT research on AARRing did not appear to have an immediately obvious way to assess relative strength using extinction procedures. One key problem is that AARRing, by definition, involves behavior that emerges and may persist in the absence of direct reinforcement for particular responses because the contingencies are extremely molar in nature. In other words, the generalized operants involved in many relational frames have relatively long reinforcement histories, going back to early language learning. Using simple extinction procedures within the context of a 1-h experimental session, for example, would not provide a realistic measure of the strength of such well-established operants. In addition, individual relational responses often form parts of larger relational networks, and thus attempting to extinguish such responses may be unsuccessful because they are maintained based on their coherence with the larger network. Admittedly, some studies on AARRing examined the extent to which it was possible to reorganize patterns of relational responding that had been established within the laboratory, such as manipulating mutually entailed relations independently of combinatorially entailed relations (e.g., Healy et al., 2000; see also Pilgrim & Galizio, 1995). Such work could thus be seen as relevant to the question of the relative strength of responding. However, this work also tended to focus on the dichotomous nature of relational frames in that it sought to establish new (reorganized) patterns that were either present or absent by the end of the training and testing procedures.

Within a few years of the publication of the Hayes et al. (2001) RFT book, therefore, the need to develop procedures that could, in principle, provide a measure of relational responding that was nondichotomous became increasingly apparent. The initial response to this need or gap in technology was the development of what came to be known as the Implicit Relational Assessment Procedure (IRAP).

The IRAP as a Measure of Relational Responding "in Flight"

The initial inspiration for the development of the IRAP was the question *How can we capture relational frames in flight?* Or, in other words, how can we measure the probability of specific patterns of AARRing, particularly those patterns that often occur in the natural environment? In developing the IRAP, two separate methodologies were combined. The first of these was an RFT-based procedure for training and testing multiple stimulus relations, the Relational Evaluation Procedure (REP), and the second was the Implicit Association Test (IAT). The latter had been developed by social cognition researchers as a method for measuring what they conceptualize as associative strengths in memory (Greenwald, McGhee, & Schwartz, 1998). When the two measures were combined into the IRAP, however, this was conceptualized as a procedure for measuring the strength or probability of AARRing (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008). Due, in part, to its close connection to the IAT, however, research with the IRAP quickly became dominated by studies focused on so-called implicit attitudes and implicit cognition more generally. Indeed, the focus on implicit cognition was likely reinforced by the relative success the IRAP achieved in this domain (see Vahey, Nicholson, & Barnes-Holmes, 2015). The story of the IRAP,

therefore, provides an excellent example of how the attempt to develop a procedure to study the dynamics of derived stimulus relations led almost inexorably to a behavior–analytic research program on human (implicit) cognition.

The IRAP: Procedural and Analytic Overview It is worth considering how the IRAP aims to provide a measure of the relative strength of derived stimulus relations. The IRAP is a computer-based task on which an individual responds to a series of screens that contain stimuli that would be defined as verbal by RFT (i.e., stimuli that have acquired their functions based, at least in part, on a history of AARRing; Hayes et al., 2001). Label stimuli—such as *flower* and *insect*—appear at the top of the screen (see Fig. 1). Target stimuli—such as *pleasant*, *good*, *unpleasant*, and *bad*—appear in the middle of the screen. On each trial, two response options are provided that specify particular relationships between the label stimuli and the target stimuli. For example, *flower* and *pleasant* might appear on a given trial with the response options *true* and *false*, and in this case participants would be required to confirm (pick *true*) or deny (pick *false*) that flowers are pleasant.

The IRAP operates by requiring opposite patterns of responding across successive blocks. For example, *flower* and *pleasant* would require the response *true* on one block and *false* on the next block. The IRAP is based on the assumption that, all things being equal, the more frequently reinforced (and thus more probable) response pattern, or one that is relationally coherent with it, will be emitted more readily (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). To increase the likelihood that the more probable response is emitted, responding on the IRAP is placed under time pressure. Within the verbal community, certain relational responses are more likely to be reinforced than punished (e.g., affirming that flowers are pleasant), whereas others are more likely to be punished than reinforced (e.g., denying that flowers are pleasant). Thus, the more readily emitted pattern of responding is indicative of the natural contingencies operating in the wider verbal community (i.e., using stimuli that are already "meaningful" to the participants). Broadly speaking, the IRAP is scored by subtracting the mean response latency for one pattern of responding from the mean response latency of the opposite pattern of responding. Any resultant difference is deemed to be reflective of the differential reinforcement for the two patterns of responding (or relationally coherent patterns) in the pre-experimental history of the individual. In most IRAP studies, four difference scores are calculated: one for each of the four trial types illustrated in Fig. 1.

Brief and Immediate Relational Responses (BIRRs) Versus Extended and Elaborated Relational Responses (EERRs) In considering the types of effects that have been obtained on the IRAP, behavioral researchers have often referred to BIRRs, which are emitted relatively quickly within a short window of time after the onset of the stimuli presented on any given IRAP trial. In contrast, EERRs are more complex and

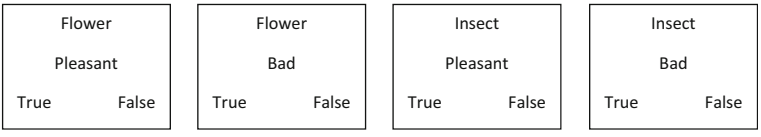

Fig. 1 An example of the four screens—or trial types—that may be presented within an Implicit Relational Assessment Procedure (IRAP)

are seen as being emitted more slowly; as such, they occur over a longer period of time (Barnes-Holmes et al., 2010; Hughes, Barnes-Holmes, & Vahey, 2012). The distinction between BIRRs and EERRs was first formalized in the context of the relational elaboration and coherence (REC) model, which was offered as an initial RFT approach to implicit cognition (Barnes-Holmes et al., 2010; Hughes et al., 2012). The basic idea behind the model is that the types of effects observed on the IRAP, and indeed other implicit measures, were because the task targeted BIRRs rather than EERRs. For example, the fact that the IRAP requires participants to respond relatively quickly on each trial almost by definition forces the participant to emit a BIRR. The relative strength or probability of this BIRR is deemed to be a function of the behavioral history of the participant with regard to functionally similar stimuli in that participant's history.

Imagine, for example, a White individual who has resided exclusively in White neighborhoods, has no non-White friends or family members, and has been exposed to many media images of Black people as violent drug dealers and inner-city gang members. When presented with an IRAP that displays pictures of Black males carrying guns, it is likely, according to the REC model, that BIRRs for confirming that Black men are "dangerous" and "criminals" may be more probable than denying such relations. As a result, the participant may respond more rapidly on the IRAP when required to confirm, rather than deny, that a Black man carrying a gun is dangerous (see Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010). In effect, an anti-Black racial bias may be revealed by the IRAP. In contrast, such a bias might be absent if the same participant were asked to rate the pictures of the Black men from the IRAP with no time constraints for doing so. The REC model predicts a lack of racial bias in the latter context by appealing to EERRs, which occur given sufficient time for an individual to respond in accordance with a relationally coherent network. In the context of the current example, the participant might fail to report any initial BIRR that involves perceiving the pictures of Black males as "dangerous" based on additional relational responding, such as *It is wrong to discriminate on the basis of race* and *I am not a racist*. In general, therefore, the REC model attempts to explain the emergence of specific response biases on the IRAP by arguing that the procedure tends to reveal BIRRs rather than EERRs.

Limitations to the REC Model: the Beginnings of a More Sophisticated Account of the IRAP In concluding that the IRAP reveals BIRRs rather than EERRs, the REC model assumes that this applies, with roughly equal force, to all four trial types (see Fig. 1). Imagine, for example, an IRAP that aimed to assess the response probabilities of four well-established verbal relations pertaining to nonvalenced stimuli such as shapes and colors. Across trials, the two label stimuli, color and shape, could be presented with target words consisting of specific colors (*red*, *green*, and *blue*) and shapes (*square*, *circle*, and *triangle*). As such, the IRAP would involve presenting four different trial types that could be designated as (a) color–color, (b) color–shape, (c) shape–color, and (d) shape–shape. During a shapes and colors IRAP, participants would be required to respond in a manner that was consistent with their pre-experimental histories during some blocks of trials: (a) color–color–true, (b) color–shape–false, (c) shape–color–false, and (d) shape–shape–true. On other blocks of trials, the participants would have to respond in a manner that was inconsistent with those histories: (a) color–color–false, (b) color–shape–true, (c) shape–color–true, and (d) shape–shape–false. Thus, when the

four trial-type effects are calculated, by subtracting response latencies for history-consistent from history-inconsistent blocks of trials, one might expect to see four roughly equal trial-type effects. In other words, the difference scores for each of the four trial types should be broadly similar. Critically, however, the patterns of trial-type difference scores obtained with the IRAP frequently differ across the four trial types (e.g., Finn, Barnes-Holmes, Hussey, & Graddy, 2016).

The REC model always allowed for the potential impact of the functions of the response options on IRAP performances. For example, Barnes-Holmes et al. (2010) pointed out that "It is possible... that a bias toward responding 'True' over 'False,' per se, interacted with the... stimulus relations presented in the IRAP" (p. 62). As such, one might expect to observe larger differences in response latencies for trial types that required a *true* rather than a *false* response during history-consistent blocks of trials. In the case of the aforementioned shapes and colors IRAP, therefore, larger IRAP effects for the color–color and shape–shape trial types might be observed relative to the remaining two trial types (i.e., color–shape and shape–color). The REC model does *not* predict that the IRAP effects for the color–color and shape–shape trial types will differ (because they both require choosing the same response option within blocks of trials), but in fact our research, both published and unpublished, has shown that they do (e.g., Finn et al., 2016, Experiment 3). Specifically, we have found what we call a "single trial type dominance effect" for the color–color trial type—that is, the size of the difference score for this trial type is often significantly larger than that for the shape–shape trial type. This finding has led us to propose an updated model of the relational responding that we typically observe on the IRAP, which we will subsequently briefly outline. A complete description of the model and its implications for research using the IRAP are beyond the scope of this article (but see Finn, Barnes-Holmes, & McEnteggart, 2017). However, we will consider the model here simply to highlight how an ongoing focus on derived stimulus relations is continuing to contribute to a behavior–analytic approach to human language and cognition (specifically the subtle variables involved in relatively simple patterns of relational responding).

In attempting to explain the single trial type dominance effect for the shapes and colors IRAP, it is first important to note that the color words we used in our research occur with relatively high frequencies in natural language in comparison with the shape words (Keuleers, Diependaele, & Brysbaert, 2010). We therefore assume that the color words evoke relatively strong orienting responses relative to the shape words (because the former occur more frequently in natural language). Or, more informally, participants may experience a type of confirmatory response to the color stimuli that is stronger than for the shape stimuli. Critically, a functionally similar confirmatory response may be likely for the *true* response option relative to the *false* response option (because *true* frequently functions as a confirmatory response in natural language). A high level of functional overlap, or what we define as coherence, thus emerges on the color–color trial type among the orienting functions of the label and target stimuli and the *true* response option. During consistent blocks, this coherence is also consistent with the relational response that is required between the label and target stimuli (e.g., color–red–true). In this sense, during consistent blocks, this trial type could be defined as involving a maximum level of coherence because all of the responses to the stimuli, both orienting and relational, are confirmatory. During inconsistent blocks, however, participants are required to choose the *false* response option, which does not cohere

with any of the other orienting or relational responses on that trial type, and this difference in coherence across blocks of trials yields relatively large difference scores. The model we have developed that aims to explain the single trial type dominance effect—and a range of other effects we have observed with the IRAP—is called the differential arbitrarily applicable relational responding effects (DAARRE) model (pronounced *dare*). We elaborate on the model in the following.

A core assumption of the DAARRE model is that differential trial-type effects may be explained by the extent to which the Cfunc and Crel properties of the stimuli contained within an IRAP cohere with specific properties of the response options across blocks of trials. Response options such as *true* and *false* are referred to as relational coherence indicators (RCIs) because they are often used to indicate the coherence or incoherence between the label and target stimuli that are presented within an IRAP (see Maloney & Barnes-Holmes, 2016, for a detailed treatment of RCIs). The basic DAARRE model as it applies to the shapes and colors IRAP is presented in Fig. 2. The model identifies three key sources of behavioral influence:

1. The relationship between the label and target stimuli (labeled Crels);
2. The orienting functions of the label and target stimuli (labeled Cfuncs); and
3. The coherence functions of the two RCIs (e.g., *true* and *false*).

Consistent with the earlier suggestion that color-related stimuli likely possess stronger orienting functions relative to shape-related stimuli (based on differential frequencies in natural language), the Cfunc property for colors is labeled positive and the Cfunc property for shapes is labeled negative. The negative labeling for shapes does not indicate a negative orienting function but simply rather an orienting function that is weaker than that of colors. The labeling of the relations between the label and target stimuli indicates the extent to which they cohere or do not cohere based on the participant's relevant history. Thus, a color–color relation is labeled with a plus sign (i.e., coherence), whereas a color–shape relation is labeled with a minus sign (i.e., incoherence). Finally, the two response options are each labeled with a plus or minus sign to indicate their functions as either coherence or incoherence indicators. In the current example, *true* (+) would typically be used in natural language to indicate coherence and *false* (−) to indicate incoherence. Note, however, that these and all of the other functions labeled in Fig. 2 are behaviorally determined by the past and current contextual history of the participant and should not be seen as absolute or inherent in the stimuli themselves.

As shown in Fig. 2, each trial type differs in its pattern of Cfuncs and Crels in terms of plus and minus properties that define the trial type for the shapes and colors IRAP. The single trial type dominance effect for the color–color trial type may be explained, as noted previously, by the DAARRE model based on the extent to which the Cfunc and Crel properties cohere with the RCI properties of the response options across blocks of trials. To appreciate this explanation, note that the Cfunc and Crel properties for the color–color trial type are all labeled with plus signs; in addition, the RCI that is deemed correct for history-consistent trials is also labeled with a plus sign (the only instance of four plus signs in the diagram). In this case, therefore, according to the model, this trial type may be considered maximally coherent during history-consistent

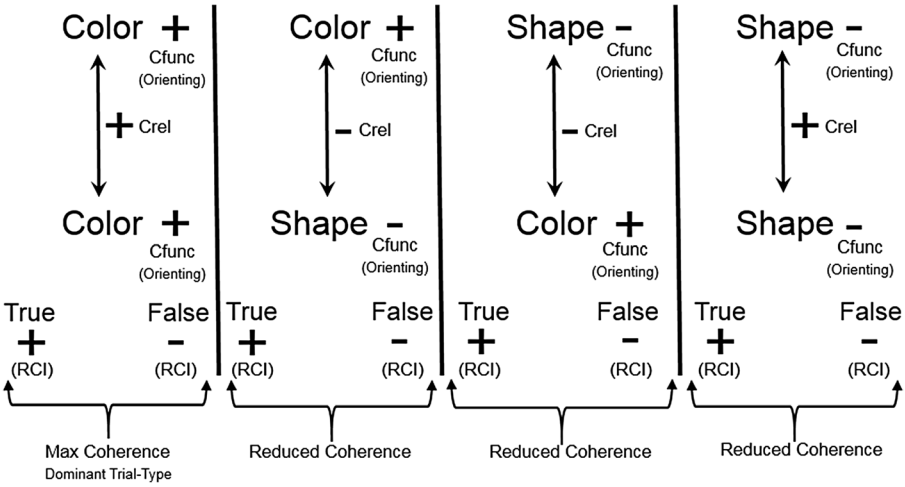

Fig. 2 The differential arbitrarily applicable relational responding effects (DAARRE) model as it applies to the shapes and colors Implicit Relational Assessment Procedure (IRAP). The positive and negative symbols refer to the relative positivity of the transformation of function property (Cfunc) for each label and target; the relative positivity of the entailment property (Crel); and the relative positivity of the relational coherence indicator (RCI) in the context of the other Cfunc, Crel, and RCI properties

trials. In contrast, during history-inconsistent trials, there is no coherence between the required RCI (minus sign) and the properties of the Cfunc and Crel (all plus signs). According to the DAARRE model, this stark contrast in levels of coherence across blocks of trials serves to produce a relatively large IRAP effect. Now consider the shape–shape trial type, which requires that participants choose the same RCI as in the color–color trial type during history-consistent trials, but here the property of the RCI (plus signs) does *not* cohere with the Cfunc properties of the label and target stimuli (both minus signs). During history-inconsistent trials, the RCI *does* cohere with the Cfunc property but not with the Crel property (plus sign). Thus, the differences in coherence between history-consistent and history-inconsistent trials across these two trial types are not equal (i.e., the difference is greater for the color–color trial type) and thus favor the single trial type dominance effect (for color–color). Finally, as can be seen in Fig. 2 for the remaining two trial types (color–shape and shape–color), the differences in coherence across history-consistent and history-inconsistent blocks are reduced relative to the color–color trial type (two plus signs relative to four), thus again supporting the single trial type dominance effect.

We are only just beginning to explore the full implications of the DAARRE model for a range of effects that we have observed with the IRAP. For example, the model may help to explain the effect of particular types of preblock instructions on IRAP performances, as reported by Finn et al. (2016), although the processes involved appear to be quite complex (see Finn et al., 2017). Furthermore, the DAARRE model becomes increasingly complex when multiple Cfunc properties (e.g., orienting vs. evaluative functions) are involved. For instance, if pictures of puppies and spiders are presented in an IRAP, the former may possess relatively strong orienting functions, but the latter may possess relatively strong (negative) evaluative functions. If and when this occurs, it complicates the model, but it makes some precise predictions that we are currently testing in laboratory studies. In any case, the main point in the current context is that

focusing on the study of derived stimulus relations across a wide range of procedures is continuing to yield increasingly sophisticated behavior–analytic treatments of many phenomena associated with human language and cognition.

Conclusion

The behavior–analytic approach to human language and cognition has been far from straightforward. The widely known critique of Skinner's (1957) *Verbal Behavior* by Chomsky (1959) almost certainly hindered the early development of a rich and vibrant program of basic research in the area. Furthermore, the absence of a focus on derived stimulus relations during the earliest stages of a child's language learning also likely contributed to the lack of impact at that time. Skinner's (1966) subsequent distinction between contingency-shaped and rule-governed behavior, however, was certainly instrumental in generating a highly active line of research that could be considered directly relevant to the study of human language and cognition. With Sidman's seminal work on equivalence relations (reviewed in Sidman, 1994) and the extension of that work into RFT during the late 1980s and 1990s, a basic behavior–analytic research program dedicated to human language and cognition began to take shape. Even today, however, some 16 years since the publication of the seminal RFT book (Hayes et al., 2001), the research program is still very much in its infancy. Relative to "mainstream" psychology, there are only a small number of active research centers, and researchers face serious challenges in attracting funding and the other resources necessary to conduct cutting-edge basic behavioral research on human language and cognition. Overall, we believe that the basic foundations have been laid, and as this article demonstrates, there is ongoing development and refinement of the theoretical and empirical analyses that have emerged in recent decades. The future clearly offers many challenges—intellectually, politically, and economically—but that future, for all its potential hurdles and difficulties, seems bright in many respects.

References

Barnes, D., Hegarty, N., & Smeets, P. M. (1997). Relating equivalence relations to equivalence relations: a relational framing model of complex human functioning. *The Analysis of Verbal Behavior, 14*, 57–83.

Barnes, D., Smeets, P. M., & Leader, G. (1996). New procedures for establishing emergent matching performances in children and adults: implications for stimulus equivalence. *Advances in Psychology, 117*, 153–171.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: two perspectives on the concept of generalized operant classes. *The Psychological Record, 50*, 251–265.

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's verbal behavior: a possible synthesis. *The Behavior Analyst, 23*, 69–84.

Barnes-Holmes, D., Barnes-Holmes, Y., Hussey, I., & Luciano, C. (2016). Relational frame theory: finding its historical and intellectual roots and reflecting upon its future development. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 115–128). West Sussex: Wiley-Blackwell.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*, 527–542.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: a preliminary study. *The Psychological Record, 58*, 497–516.

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57–80.

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: Plenum.

Barnes-Holmes, Y. (2001). *Analysing relational frames: studying language and cognition in young children (unpublished doctoral thesis)*. Maynooth: National University of Ireland.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy, 4*, 531–558.

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*, 165–180.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: theory and applications* (2nd ed.). New York: Springer.

Catania, A. C., Shimoff, E., & Matthews, B. A. (1989). An experimental analysis of rule-governed behavior. In S. C. Hayes (Ed.), *Rule-governed behavior: cognition, contingencies, and instructional control* (pp. 119–150). New York: Plenum.

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language, 35*, 26–58.

De Houwer, J. (2017). A functional-cognitive framework for cooperation between functional and cognitive researchers in the context of stimulus relations research. *The Behavior Analyst*. Advance online publication. doi:10.1007/s40614-017-0089-6.

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior, 88*, 179–197.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior, 64*, 163–184.

Dymond, S., Roche, B., & Barnes-Holmes, D. (2003). The continuity strategy, human behavior, and behavior analysis. *The Psychological Record, 53*, 333–347.

Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: the impact of three types of introductory rules. *The Psychological Record, 66*, 309–321.

Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2017). Exploring the single-trial-type-dominance-effect on the IRAP: developing a Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model. Manuscript submitted for publication.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Hayes, S. C. (1984). Making sense of spirituality. *Behavior, 12*, 99–110.

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

Hayes, S. C. (1991). *A relational control theory of stimulus equivalence*. Reno: Context Press.

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

Hayes, S. C., Gifford, E. V., Townsend, R. C., & Barnes-Holmes, D. (2001). Thinking, problem-solving, and pragmatic verbal analysis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: a post-Skinnerian account of human language and cognition* (pp. 87–101). New York: Plenum.

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior, 74*, 207–227.

Hughes, S., & Barnes-Holmes, D. (2016a). 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: Wiley-Blackwell.

Hughes, S., & Barnes-Holmes, D. (2016b). Relational frame theory: implications for the study of human language and cognition. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 179–226). West Sussex: Wiley-Blackwell.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: a voyage through implicit cognition research. *Journal of Contextual Behavioral Science, 1*, 17–38.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1*, 174.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56*, 201–239.

Lowe, C. F. (1979). Determinants of human operant behavior. *Advances in Analysis of Behaviour, 1*, 159–192.

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.

Luciano, C., Becerra, I. G., & Valverde, M. R. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior, 87*, 349–365.

Maloney, E., & Barnes-Holmes, D. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: the role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record, 66*, 395–403.

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: a developmental profile. *The Psychological Record, 54*, 115–144.

McKeel, A. N., Rowsey, K. E., Belisle, J., Dixon, M. R., & Szekely, S. (2015). Teaching complex verbal operants with the PEAK relational training system. *Behavior Analysis in Practice, 8*, 241–244. doi:10.1007/s40617-015-0067-y.

Moore, J. (2009). Some thoughts on the relation between derived relational responding and verbal behavior. *European Journal of Behavior Analysis, 10*, 31–47.

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

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63*, 225–238.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67*, 275–301.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research, 14*, 5–13.

Sidman, M. (1994). *Stimulus equivalence: a research story*. Boston: Authors Cooperative.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1966). An operant analysis of problem-solving. In B. Kleinmuntz (Ed.), *Problem solving: research, method, teaching* (pp. 225–257). New York: Wiley.

Stewart, I., & Barnes-Holmes, D. (2004). Relational frame theory and analogical reasoning: empirical investigations. *International Journal of Psychology and Psychological Therapy, 4*, 241–262.

Stewart, I., Barnes-Holmes, D., Hayes, S. C., & Lipkens, R. (2001). In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: a post-Skinnerian account of human language and cognition* (pp. 73–86). New York: Plenum.

Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 59–65.

Vaughan, M. (1989). Rule-governed behavior in behavior analysis. In S. C. Hayes (Ed.), *Rule-governed behavior: cognition, contingencies, and instructional control* (pp. 97–118). New York: Plenum.

Weiner, H. (1969). Conditioning history and the control of human avoidance and escape responding. *Journal of the Experimental Analysis of Behavior, 12*, 1039–1043.