Updating RFT (More Field than Frame) and its Implications for Process-based Therapy

Dermot Barnes-Holmes, Yvonne Barnes-Holmes, and Ciara McEnteggart

Department of Experimental, Clinical, and Health Psychology

Ghent University

Corresponding Author:

Dr. Ciara McEnteggart

Department of Experimental, Clinical, and Health Psychology

Ghent University

Henri Dunantlaan, 2

B-9000 Ghent

Belgium

Email: Ciara.McEnteggart@ugent.be

Authors' Note This article was prepared with the support of an Odysseus Group 1 grant awarded to the first author by the Flanders Science Foundation (FWO). Correspondence concerning this article should be sent to Ciara.McEnteggart@ugent.be The first part of the article was based on material provided primarily by the first author, while the second part was provided largely by the second and third authors.

Abstract

The current article presents a response to the recent call for a focus on psychological processes of change in psychotherapy. In addressing the need for a more process-based focus, the need for clarity in defining psychological processes per se becomes apparent, before it is possible to develop process-based therapy. In grappling with this challenge, the current article is divided into two parts. In Part 1, we present a modern view of behavioral processes as they apply specifically to verbally-sophisticated humans. The view we offer is based on one of the main approaches to human language and cognition within behavioral science, Relational Frame Theory (RFT), which has been updated in recent years. In Part 2, the view of behavioral processes, as seen through the lens of an updated RFT, is used to begin to develop a process-based approach to the assessment and treatment of human psychological suffering. The article ends with two case summaries and a series of brief take-home messages that aim to capture the core elements of the RFT-driven process-based therapy we are currently developing.

Key words: RFT, process-based therapy, HDML framework, ROE

In recent times, there has been a call to focus on psychological processes of change in cognitive behavior therapy (CBT) and in psychotherapy more generally (Hoffman & Hayes, 2019). The motivation behind this call is to progress beyond a strictly syndrome-based approach to human psychopathology (hereafter referred to as psychological suffering). There are two expected broad outcomes from this new focus on processes rather than syndromes. First, it will allow researchers and clinicians to develop and refine their understanding of the psychological processes inherent in psychological suffering. Second, it will facilitate improvements in both clinical assessment and treatment, because the fundamental processes may be targeted with increased precision.

We certainly see the value in calling for a greater focus on the processes involved in psychological suffering. But, of course, this immediately begs the question -- what are these psychological processes? Answering this question is far from a simple matter. There are deep divisions within psychology as a whole and these are manifested, to some extent, within clinical psychology (Wilson, 2012). Perhaps one of the most salient historical divides has been between the behavioral and cognitive traditions (Miller, 2003). The primary goal of the former is to identify and analyze the *behavioral processes* involved in the interactions that occur between the organism and its environment. The primary goal of the latter is to identify and analyze the *mental processes* involved in environment-behavior interactions (Bargh & Ferguson, 2000). Needless to say, therefore, these two traditions differ fundamentally in their views of what constitutes a psychological process, which of course complicates any attempt to develop process-based psychotherapy. Or to put it another way, developing a more process-based approach to any domain in psychology will be extremely challenging if the field as a whole cannot agree on what constitutes a psychological process. It is not our intention to resolve this issue in the current article, because the lack of agreement, in our view, is based largely on philosophical differences between the two (behavioral and cognitive) traditions (e.g., Barnes & Holmes, 1991; Hayes & Brownstein, 1986). We also do not intend to examine or debate these philosophical issues here. Rather our primary aim is to consider the implications of the call for process-based psychotherapy from *within* a modern behavioral approach to human language and cognition.

The core message that we hope to deliver in the current article, we suspect, could be seen as challenging and indeed controversial. On balance, we do not believe that our message is hasty or unwarranted because it is based on concepts and research, the roots of which may be traced back to almost 50 years ago. Of course, the message may in time turn out to be more or less useful in facilitating prediction-and-influence with precision, scope, and depth, but the empirical evidence upon which it is based cannot be ignored by either the behavioral or cognitive traditions.

The current article is divided into two parts. In Part 1, we present a modern view of behavioral processes as they apply specifically to verbally-sophisticated humans. The view we offer is based on one of the main approaches to human language and cognition within behavioral science, Relational Frame Theory (RFT, see Hayes, Barnes-Holmes, & Roche, 2001), which has been updated in recent years[1] (Barnes-Holmes, Barnes-Holmes, Luciano, & McEnteggart, 2017; Barnes-Holmes, Finn, McEnteggart, & Barnes-Holmes, 2018; Finn, Barnes-Holmes, & McEnteggart, 2018). Much of the material presented in Part 1 is also contained in a recent article (Barnes-Holmes, Finn et al.), but is repeated here to provide the full context for Part 2. In Part 2, the view of behavioral processes, as seen through the lens of

[1] The use of the term "up-dated" is employed here to reflect the fact that the current article proposes a number of new concepts for RFT that do not appear in the seminal volume by Hayes et al. (2001). It is also worth noting that the authors of the current article are not alone in "up-dating" RFT. For example, Hayes, Sanford, and Chin (2018) have recently proposed an up-dated view of how RFT connects with evolutionary science, with a focus on the role of cooperation in the evolution of AARRing itself (Hayes & Sanford, 2014). Furthermore, a number of new RFT-based concepts have also been proposed in a recent volume on the clinical application of RFT (Villatte, Villatte, & Hayes, 2015).

an updated RFT, is used to begin to develop a process-based approach to the assessment and treatment of human psychological suffering.

Part 1

Behavioral Processes as Seen through the Lens of an Updated RFT

"There is a difference between behaving and reporting that one is behaving, or reporting the causes of one's behavior. In arranging conditions under which a person describes the public or private world in which he lives, a community generates that very special form of behavior called knowing." (Skinner, 1974, p. 30).

The opening quotation to Part 1 could be seen as arguing that there is something special or fundamental about *human* knowing or "self-awareness" (see Dymond & Barnes, 1997, for a detailed discussion). Although empirical research has shown that nonhumans demonstrate self-discrimination (e.g., Lattal, 1975; Shimp, 1982), and thus "knowing" in the sense described by Skinner may not be uniquely human, *derived* self-discrimination (as articulated within RFT) has also been shown in humans (e.g., Dymond & Barnes, 1994, 1995). Insofar as such derived effects are largely restricted to the human species (see Hughes & Barnes-Holmes, 2014, for a recent discussion), it could be argued that "*derived* knowing" is the *very special form of behavior* to which Skinner was referring.

Of course, it is now over 40 years since Skinner wrote those words and thus we are left only with idle speculation when interpreting the exact meaning of the opening quotation. In our view, however, we suspect that Skinner may have been pointing to a key feature of human behavior that requires a careful and systematic analysis in its own right. Such an analysis has, in a sense, been unfolding since Skinner (1966) proposed the concept of rule-governed behavior, but it is not our intention to cover that history here because it has been articulated elsewhere (see Barnes-Holmes, Finn et al., 2018). The important point is that the study of human *knowing*, or perhaps more accurately *derived knowing*, could be seen as lying at the cutting edge of RFT. Part 1 of the current article is focused on this recent work.

Conceptual Background

The key purpose of Part 1 of the current article is to integrate two recent conceptual developments within RFT. The first of these is the multi-dimensional, multi-level (MDML) framework (see Barnes-Holmes et al., 2017, for a detailed treatment) and the second is the differential arbitrarily applicable relational responding effects (DAARRE; pronounced "dare") model (e.g., Finn et al., 2018). Both the MDML framework and the DAARRE model were outlined in a recent article on the role of derived stimulus relations in behavior analytic research on human language and cognition (Barnes-Holmes, Finn et al., 2018), but no attempt was made to integrate these two recent developments into a single conceptual framework.

An integration of the MDML framework and the DAARRE model is offered here because doing so appears to capture the two core defining properties of arbitrarily applicable relational responding (AARRing) itself, entailment relations and the derived transformation of functions.[2] It should be emphasized at the outset that these two properties were contained in both the MDML framework and the DAARRE model, as originally proposed, but particularly in the case of the MDML framework the emphasis appeared to be on entailment alone. The current article attempts to place a greater emphasis on function transformation within the MDML framework by drawing on the DAARRE model, thus yielding what we now call a *hyper*-dimensional, multi-level (HDML) framework for analyzing the dynamics of AARRing. Before continuing, it is important to note that we will aim to provide sufficient background material to appreciate what may be gained through the proposed conceptual integration, but it would be useful for a reader who has not previously encountered the MDML framework or the DAARRE model to first read the recent article by Barnes-Holmes, Finn et al. (2018).

Relational Frame Theory

---

[2] Relational frames have been defined as consisting of three properties; mutual entailment, combinatorial entailment, and the transformation of functions, which we are not challenging here. As will become apparent, however, the MDML framework is focused on arbitrarily applicable relational responding in general, not relational frames specifically. In the context of the MDML framework, therefore, it seems wise to refer to two general properties of AARRing, entailment and transformation of functions.

Relational frame theory emerged directly from the seminal work of Murray Sidman on what he called *stimulus equivalence* (e.g., Sidman, 1971, 1994; Sidman & Tailby, 1982). 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., A-B and A-C), untrained matching responses frequently appeared in the absence of additional learning (e.g., B-C and C-B). 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.

Relational frame theory appeared to provide an important extension to the work on stimulus equivalence (see Hughes & Barnes-Holmes, 2016a, 2016b, for recent extensive reviews). Specifically, the theory argued that stimulus equivalence may be considered a generalized relational operant, and 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 cat upon hearing the word “cat” or the specific cat’s name (e.g., “Tibbles”), and to emit other appropriate naming responses, such as saying “Tibbles” or “cat” when the family pet was observed, or saying “Tibbles” when asked, “What is the cat’s name?” Across many such exemplars, involving other stimuli and contexts, eventually the operant class of coordinating stimuli in this way is established, such that direct reinforcement for all of the individual components of naming are no longer required when novel stimuli are encountered. Imagine, for example, that the child was shown a picture of a platypus, and the written word, and was told its name. Subsequently, the child may say “That’s a platypus” 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, 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 AARRing has been established, the generalized relational responding may then be applied to any stimuli, given appropriate contextual cues.[3]

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 above. Thus, for example, the phrase “that is a”, as in “*That is a* cat”, would be established across exemplars as a contextual cue for the complete pattern of relational responding (e.g., coordinating the word “cat” with actual cats). Similarly, phrases such as “That is not a”, or “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 produced in any instance in which stimuli are related. For example, the word “cat” could produce different responses for a child if she was asked “what does your cat look like” versus “feel like”. In RFT, therefore, two broad classes of contextual cues are involved

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

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 produced during the act of relating; these two classes of contextual cue 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 A is related to C, then B is related to C, and C is related to B), 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 *more than* B and B was *more than* C in a particular instance of relational framing, and a reinforcing function was attached to A, then C may acquire a reduced reinforcing function relative to A or B. The concept of a relational frame was thus designed to capture how human language and cognition change 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 AARRing. Even a simple tendency to orient strongly towards a particular stimulus in your visual field may be based on AARRing. For example, a startle response to the movement of a spider may occur more quickly or strongly in a foreign country because a native speaker recently coordinated local "spiders" with "highly venomous." Such functions may be defined as Cfunc properties because they are examples of specific stimulus functions (i.e., orienting/startle) that are acquired based on, but are separate from, the entailed relations among the relevant stimuli.[4]

Much of the early empirical work in RFT focused on showing that AARRing may 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). 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, studies showed that AARRing could be observed using a variety of procedures (e.g., Barnes, Smeets, & Leader, 1996), indicating that the phenomenon was not tied to a particular experimental preparation or modes of instruction, providing 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

[4] The reader should note that within RFT Crel and Cfunc properties are not separable units of analysis but separate properties of the single unit (i.e., the relational frame). In the scientific act of any given experimental, applied, or conceptual analysis, a greater or lesser focus may be targeted on the Crel or Cfunc properties of a particular pattern of AARRing. However, it would be a mistake to think that a Cfunc property may be isolated meaningfully from a Crel property, or vice versa. Indeed, as pointed out by Dymond and Barnes (1994, p. 264) a quarter of a century ago, "the relational-frame account. . . views. . . equivalence responding and derived transfer of function. . . as products of the single behavioral process of arbitrarily applicable relational responding. . . In effect, the observed pattern of a transfer of functions defines the entailed relations, and thus the entailed relations. . . do not exist as a behavioral event until a specific transfer of functions has occurred." Or to put it another way, whenever a Cfunc property of a stimulus is identified in a particular analysis it must be defined in terms of a particular Crel property. That is, virtually all psychological acts for verbal humans involve the process of entailment. We shall return to this issue later in the current article.

traction (see Barnes-Holmes & Barnes-Holmes, 2000; Healy, Barnes-Holmes, & Smeets, 2000).

The seminal text on RFT (Hayes, et al., 2001) also used the basic operant unit of the relational frame to scale up to more advanced levels of AARRing when analyzing behavior in specific domains of human language and cognition (see also Hayes & Hayes, 1989). For example, rule-governed behavior was interpreted as a network of relational frames involving coordination and temporal relations, and the necessary contextual cues, to transform specific behavioral functions (Barnes-Holmes, O'Hora et al., 2001). Take the 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. Empirical models of verbal rules or instructions as complex relational networks composed of multiple relational frames have since been published (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004; O'Hora, Barnes-Holmes, & Stewart, 2014).

Scaling up to even more advanced levels in RFT was involved in analyzing behavior in other areas of human language and cognition, such as analogical and metaphorical reasoning (e.g., Barnes, Hegarty, & Smeets, 1997). 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 inherent in relational framing more generally (i.e., AARRing), but applied to framing itself (see Stewart & Barnes-Holmes, 2004, for a review of empirical research involving both adults and young children). An even more advanced level of AARRing is proposed by RFT, which involves relating complex relational networks to other complex relational networks. Although research in this area is extremely limited (cf. Ruiz & Luciano, 2011), advanced verbal abilities, such as those involved in comparing and contrasting extended narratives, would appear to require this level of AARRing (see Barnes-Holmes, Barnes-Holmes, et al., 2017).

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[5]: the interpersonal relations I-YOU, spatial relations HERE-THERE, and 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 they are eating an evening meal with their family. If the child responded simply by referring to what a sibling is currently having for dinner, they 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 on-going refinement of the three deictic relations allows the child to respond appropriately to questions about their own behavior in relation to others, as it occurs in specific times and specific places (e.g., McHugh, Barnes-Holmes, & Barnes-Holmes, 2004).

Although RFT highlights the three deictic relations (I-YOU; HERE-THERE, and NOW-THEN) in explaining the emergence of a *verbal self*, the foregoing examples illustrate that establishing such a self is very much entangled with increasingly complex AARRing. For example, learning to respond appropriately to questions about eating lunch versus dinner likely requires relating relational networks without the added complexity of specifying who

[5] 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 time (e.g., "now" rather than "then").

was eating which food and at which time. In other words, the extent to which the verbal community can establish a verbal self very much depends on establishing increasingly advanced forms of AARRing; and of course as a verbal self emerges this likely allows the verbal community to develop and refine increasingly advanced forms of AARRing in general. That is, the relationship between advanced forms of AARRing and the verbal self should be seen as non-linear and dynamic. Indeed, this feature of RFT cannot be over-emphasized, and lies at the very core of our interpretation of that very special form of behavior that Skinner defined as *knowing*. Or to put it another way, without AARRing there would be no derived knowing or verbal self, and without derived knowing (or a verbal self) AARRing, at best, would be extremely constrained and limited, and would not allow for the richness, complexity, and power of human language and cognition.

The MDML Framework

The previous sections provided a very brief overview of how RFT has been used to provide functional-analytic accounts of numerous areas of human language and cognition (see Hughes & Barnes-Holmes, 2016b, for a recent review). In an effort to systematize the RFT account, researchers have recently offered the MDML as a framework for analyzing AARRing (Barnes-Holmes, Barnes-Holmes, et al., 2017; see Table 1). According to this framework, AARRing may be conceptualized as developing in a broad sense from; (i) mutual entailing, to (ii) simple networking involved in frames, to (iii) more complex networking involved in rules and instructions, to (iv) the relating of relations involved in analogical reasoning, and finally to (v) relating relational networks, which is typically involved in understanding and producing extended narratives, and advanced problem-solving. In identifying these as different levels, the MDML framework is not indicating that they are rigid or invariant *stages*. Rather, lower levels 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. 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.

Insert Table 1 Here

A brief description of each of the four dimensions is as follows. *Coherence* refers to the extent to which specific patterns of AARRing are generally consistent with other patterns of AARRing. For example, the statement “A car is larger than a truck” 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 involves two types of relations (if A is *bigger than* B, then B is *smaller than* A).[6] *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

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

asked to respond with the wrong answer to the question “Which is bigger, a car or a truck?” The easier this is achieved, the more flexible the AARRing.

In a sense, the MDML framework simply makes explicit what basic researchers in RFT have been doing implicitly since the theory was first subjected to experimental analysis. That is, whenever an RFT researcher conducts a lab-based study it 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 specifies how many trials will be used to test for the entailed symmetry relations (e.g., 12), and how many trials must be *correct* to define the performance as mutual entailment (e.g., 10/12). In effect, the number of opportunities to *derive* the entailed relations must be specified (i.e., 12) and the number of responses that must *cohere* with the relations is also determined (i.e., 10). In effect, 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 changing the baseline training), then relational *flexibility* in the original test performance is also assessed. As noted above, RFT and equivalence researchers 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 AARRing identified in RFT and the well-established dimensions along which the levels have been, or could be, studied.

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 above, the tested relation must *cohere* in some pre-specified manner with the trained relation (e.g., if A is taller than B, then B will be shorter than A), and the number of *derived* relational responses must be specified (e.g., a participant must produce at least 10 out of 12 responses indicating that B is indeed shorter than A in the absence of programmed reinforcement, prompting or other feedback).

A detailed treatment of the MDML framework has been provided elsewhere (e.g., Barnes-Holmes, Barnes-Holmes, et al., 2017) and thus there is no need to work through all the details and subtleties here. The critical point is that RFT may be used to generate a conceptual framework that begins with a basic scientific unit of analysis, the mutually entailed derived stimulus relation, identifying at least some of the key dimensions along which mutual entailment may vary (e.g., 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 dimensions listed above, and perhaps others that remain to be identified.

The DAARRE Model

As noted at the beginning of the current article, the MDML framework appears to be very much focused on the entailment relations (or Crel properties) of AARRing. The function-transformation (or Cfunc properties) of AARRing has always been assumed within the MDML framework, because RFT defines AARRing itself in terms of both properties. Nevertheless, it seems important to incorporate these properties into the MDML framework in a relatively explicit manner. This objective, we believe, may be achieved by integrating the MDML framework with another recent development in RFT, the DAARRE model.

The DAARRE model emerged primarily from research conducted using the Implicit Relational Assessment Procedure (IRAP), a methodology which is based on RFT itself. We will not present a detailed treatment of that method here because relevant material is available in many other published sources (but see Barnes-Holmes, Finn et al., 2018, for a recent summary). We will, however, provide sufficient detail in what follows so that the reader may appreciate the close connection of the DAARRE model to the IRAP.

The IRAP was developed initially as a method for assessing the strength or probability of verbal relations in natural language, as conceptualized by RFT (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008). For illustrative purposes, consider an IRAP that aimed to assess the response probabilities of four well-established verbal relations pertaining to non-valenced 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 (i) *Color-Color,* (ii) *Color-Shape,* (iii) *Shape-Color,* and (iv) *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; (i) *Color-Color-True;* (ii) *Color-Shape-False;* (iii) *Shape-Color-False;* and (iv) *Shape-Shape-True.* On other blocks of trials, the participants would have to respond in a manner that was inconsistent with those histories; (i) *Color-Color-False;* (ii) *Color-Shape-True;* (iii) *Shape-Color-True;* and (iv) *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 pattern of trial-type difference-scores obtained with the IRAP frequently differ across the four trial-types (e.g., Finn, Barnes-Holmes, Hussey, & Graddy, 2016).

Early research with the IRAP always allowed for the potential impact of the functions of the response options on IRAP performances. For example, Barnes-Holmes, Murphy, Barnes-Holmes, and Stewart (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 "Shapes-and-Colors" IRAP described above, 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*). Of course, this analysis 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 for the *Shape-Shape* trial-type. This finding led us to propose the DAARRE model of the response patterns that are typically observed on the IRAP, which we will briefly outline subsequently (a complete description of the model, and its implications for research using the IRAP, is beyond the scope of the current article; but see Finn et al., 2018; see also, Kavanagh, Barnes-Holmes, Barnes-Holmes, McEnteggart, & Finn, 2018).

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 elicit 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” relative to the “False” response option (because “True” frequently functions as a confirmatory response in natural language). A high level of functional overlap, or 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 (among the orienting functions) also coheres with the relational response (or Crel property) 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.

A core assumption of the DAARRE model, therefore, 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. The reader should note that 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 Figure 1. The model identifies three key sources of behavioral influence: (1) the relationship between the label and target stimuli (labeled as Crels); (2) the orienting functions of the label and target stimuli (labeled as 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 as positive and the Cfunc property for Shapes is labeled as negative. The negative labeling for shapes should not to be taken to indicate a negative orienting function but simply 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 participants' 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 Figure 1 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.

Insert Figure 1 Here

As noted above, the single-trial-type-dominance-effect for the *Color-Color* trial-type may be explained 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 as maximally coherent during history-consistent trials. In contrast, during history-inconsistent trials there is no coherence between the required RCI (minus sign) and the properties of the Cfuncs 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 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 properties 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 is not equal (i.e., the difference is greater for the *Color-Color* trial-type) and thus favors the single-trial-type-dominance-effect (for *Color-Color*). Finally, as becomes apparent from inspecting Figure 1 for the remaining two trial-types (*Color-Shape* and *Shape-Color*) the differences in coherence across history-consistent and history-inconsistent blocks is reduced relative to the *Color-Color* trial-type (two plus signs relative to four), thus again supporting the single-trial-type-dominance-effect.

The DAARRE model becomes increasingly complex when multiple Cfunc properties are involved. Consider, for example, that we not only notice or orient toward specific stimuli or events, but we also may react to those stimuli as relatively appetitive or aversive (defined here as *evoking* functions). For illustrative purposes, imagine that instead of using words referring to shapes and colors we inserted pictures of cute and cuddly puppies or kittens into an IRAP as one category of stimuli, along with pictures of large and aggressive-looking spiders as the other category. Imagine also that participants were required to respond to these pictures with either *approach* (e.g., “I can pick it up”) or *avoidance* (e.g., “I need to get away”) descriptors (for recent research using these types of IRAPs, see Leech, Barnes-Holmes, & Madden, 2016; Leech, Barnes-Holmes, & McEnteggart, 2017). In this case, it seems likely that two separate Cfunc properties (i.e., orienting and evoking) could play a role in determining responding on the IRAP. For example, the pictures of spiders, as potentially dangerous or threatening stimuli, may possess relatively strong orienting and aversive evoking functions, relative to the pictures of pets. Indeed, the latter, as cute and cuddly stimuli, would likely possess relatively strong appetitive evoking functions (but perhaps relatively weaker orienting functions due to their lack of threat/danger). The approach and avoidance descriptors may not possess orienting functions that differ dramatically from each other, but it seems likely that they would differ in terms of evoking functions (i.e., approach = appetitive and avoidance = aversive). For illustrative purposes, the DAARRE model interpretation for one of the trial-types (*Spider-Approach*) is presented in Figure 2. We present this trial-type in particular because it has yielded potentially interesting effects in the two recently published studies by Leech, et al., to which we now turn.

Insert Figure 2 Here

Specifically, the *Spider-Approach* trial-type has tended to produce an IRAP effect that is opposite in direction (or extremely weak) to that which might be predicted based on the assumption that participants would not readily approach spiders in the natural environment. In other words, across the two studies there was a small response bias toward pressing “True” more quickly than “False.” Critically, however, the response biases on this particular trial-type were significantly correlated with participants’ performances on a behavioral approach task involving a live spider. That is, a stronger tendency to respond more quickly with “True” than with “False” on the *Spider-Approach* trial-type was associated with increased levels of approach towards an actual spider. Thus, although the direction of the IRAP effect could be seen as *counter-intuitive*, it predicted actual behavior. Although entirely post-hoc and somewhat speculative, the DAARRE model may be used to explain this outcome. Let us assume that for participants who were relatively low in spider fear, the orienting function of spiders on the IRAP dominated over the evoking function, because the latter (function) was not particularly aversive or appetitive. However, for participants who were relatively high in fear the (aversive) evoking function dominated over the orienting function.[7] If this was the case, then responding "True" would be more coherent than responding "False" for low-fear participants, whereas this would not be the case for high-fear participants. To appreciate the argument we are making consider Figure 2.

The figure indicates that there is a negative Crel between spiders and approach (i.e., most people would report that do not readily approach spiders). Thus, a *correct* response on a history-consistent trial would be "False." However, the wider context of the IRAP establishes a relatively strong spider orienting function for participants who are low in spider-fear, but a relatively strong aversive evoking function for participants who are high in spider-fear. For the low-fear individuals, therefore, the dominating Cfunc for spiders (orienting) is positive as is the evoking Cfunc for the approach descriptor, both of which cohere with the positivity of the "True" RCI. More informally, low-fear participants may experience a type of "Yes-Yes" effect when presented with this trial-type in an IRAP, which results in a tendency to pick "True" more quickly than "False." For the high-fear individuals, however, the dominating Cfunc for spiders (evoking) is negative but positive for the approach descriptor, and thus one of the Cfuncs coheres with the "True" RCI and the other coheres with the "False" RCI. More informally, high-fear participants may experience a type of "No-Yes" effect with this trial-

[7] The participants in the studies reported by Leech et al. (2016, 2017) were recruited randomly from normative samples and thus were not formally categorized as high and low in levels of self-reported fear of spiders or in their tendency to approach actual spiders. Nevertheless, self-reported fear, and performance on a behavioral approach task, were found to vary within the sample, and thus at least some evidence of a correlation between the IRAP performances and behavioral approach might be expected.

type, which reduces the tendency to pick "True" over "False", at least when compared to the low-fear participants.

As noted above, the DAARRE model interpretation offered here is post-hoc and speculative, but critically it serves to highlight the potential complexities involved in analyzing and explaining an IRAP performance when the Crel property between the label and target stimuli is balanced against the impact of multiple Cfunc properties for those stimuli.[8] Or in other words, most participants may entail a *not* relation between spiders and approach but differ in the extent to which they confirm or disconfirm this relation in the context of an IRAP, based on the relative dominance of Cfunc orienting and evoking functions for the individual stimuli and the RCIs. Indeed, recognizing this complexity has led us to integrate the MDML framework with the DAARRE model, yielding a hyper-dimensional multi-level framework (or HDML).

In proposing the HDML framework, the conceptual and empirical analyses suggested by RFT encourage focusing on a *field of interactants*, rather than on individual frames. For example, one only need examine Figure 1 to appreciate the extent to which explaining performance on a simple IRAP requires a functional analysis of the interacting nature of the Crel and Cfunc properties. Indeed, it is worth noting that a strong Kantorian (i.e., field-like) influence on the initial formulation of RFT was provided by L. J. Hayes (see prologue of Hayes et al., 2001, p. viii). In shifting the emphasis from individual frames to relational networks or fields of interactants, we propose a new conceptual unit of analysis for RFT, which we refer to as the ROE (pronounced "row"), which stands for relating, orienting, and

---

[8] It is important to note that the DAARRE model remains a work in progress, and as such is being used to interpret a range of effects that have been observed in IRAP performances (e.g., Kavanagh, Roelandt, Van Raemdonck, Barnes-Holmes, Barnes-Holmes, & McEnteggart, in press). We have not covered this work here, however, because the key point has been made – the AARRing involved in relatively simple relational networks, even when responses must occur within relatively brief periods of time (i.e., in an IRAP), typically involves a complex cluster of controlling variables.

evoking. In the next section of the article we will consider the proposed integration, the ROE, and some of the potential implications arising therefrom.

Integrating the MDML Framework with the DAARRE Model: Relating, Orienting, and Evoking (ROEing) as a Conceptual Unit of Analysis

At this point, it should be clear that completing an IRAP involves a dynamic interplay among the Crel and Cfunc properties of the stimuli presented within the procedure. Insofar as an IRAP may provide laboratory analogues of the types of relational networking (or AARRing) that occur in the natural environment, the systematic functional analyses of IRAP performances may yield important insights into the controlling variables that are at play as verbally-able humans navigate their public and private worlds. Certainly, some elements of the potential complexity involved in analyzing these variables were highlighted in the MDML framework. Upon reflection, however, the property of entailment appears to be the key focus of the MDML framework, but as the foregoing material on the DAARRE model highlights the dynamics involved in AARRing also involve focusing on the transformation of functions (or Cfunc properties). Recently, therefore, we have updated the MDML framework by integrating it with the DAARRE model, and now refer to the MDML as the *H*DML, where the ‘H’ stands for ‘hyper’. The term “hyper” is used because the integration does not simply involve adding additional dimensions to the MDML framework but adding new foci. At the present time, these new foci are the orienting and evoking functions of the stimuli that are involved in the patterns of AARRing identified within the original MDML framework. The integration of the MDML framework and the DAARRE model is represented in Figure 3.

Insert Figure 3 Here

As can be seen in the figure, the orienting and evoking functions are represented with an inverted ‘T’ shape being placed into each of the 20 intersections of the HDML framework. The vertical line represents the relative value of orienting functions from low to high, with 0 representing the absence of any orienting function and 1 representing some pre-defined maximum value for the function (e.g., when an orienting response occurs with a probability of 1.0). The horizontal line in the inverted 'T' represents the relative value of evoking functions from extremely aversive (on the far left) to extremely appetitive (on the far right). The most aversive and appetitive functions are represented by the values of -1 and +1, respectively (e.g., if an aversive reaction occurred to a stimulus with a probability of 1.0 then the value assigned to the function would be -1, whereas if an appetitive reaction occurred with a probability of 1.0 the value assigned to the function would be +1). Both orienting and evoking functions may impact upon, and may be impacted by, the relational or entailment properties represented within each of the 20 units of the HDML framework. And virtually any contextual variable may be involved in influencing the dynamical interplay among the three properties within or across cells.

In recognizing this dynamic interplay, it seems useful to conceptualize psychological events for verbally-able humans as involving a constant behavioral stream of relating (R), orienting (O), and evoking (E), summarized as ROEing.[9] In brief, *relating* refers to the myriad complex ways in which stimuli or events may be related verbally; *orienting* refers to noticing or attending to a stimulus or event; and *evoking* refers to whether a noticed stimulus or event is appetitive, aversive, or relatively neutral. The three elements of the ROE are conceptualized as working together, synergistically, in virtually every behavioral event for a verbally-able human. For illustrative purposes, imagine you are about to enter a forest with a tour guide who warns you, "Watch out for black spiders with a red triangle on the back because they are

[9] The ROE is a new and relatively broad conceptual unit of analysis within RFT. For example, the ROE is clearly broader than the concept of a relational frame, in that it aims to capture the most basic to the most complex patterns of AARRing from mutual entailing, to framing, to complex relational networking, to relating relations, and finally to relating relational networks. The concept of the ROE may thus encourage conceptual analyses that extend beyond the level of the frame and also encourage analyses that explicitly consider the role played by the Crel and Cfunc properties of the stimuli or events that participate in any given instance of AARRing. The potential benefits of encouraging these broader types of conceptual analyses, while remaining closely linked to experimental and applied analyses, will be illustrated in Part 2 of the current article.

quite aggressive and also highly venomous." If the warning is understood, it may be conceptualized as involving an instance of relating (e.g., relating spiders with particular properties to danger), which may increase the likelihood that you will *orient* towards any spider-like shape or movement in the forest followed by an appropriate *evoked* reaction, such as backing away, freezing, or swatting it away with a stick if the object is perceived to be a black spider with a red triangle on its back. In effect, your reaction to the spider in the forest is conceptualized as involving the three elements of the ROE.

As noted above, the three elements of the ROE are not seen as interacting in a linear or unidirectional manner, but are dynamical. Thus, for example, an orienting response may produce relating, which then leads to an evoked response. Imagine you entered the forest without hearing any warning about spiders. You might be less likely to orient toward spider-like movements, in the absence of the previous warning, but if you did notice a spider you may engage in some relational activity, such as emitting the self-generated rule "better safe than sorry" and withdrawing slowly. In this latter case, orienting led to relating, which led to evoking.

Before proceeding, it is important to stress, as noted above, that contextual variables, such as motivational factors, will participate in the field of interactants that influence the dynamical interplay among the three properties of the ROE, within or across cells. For instance, water deprivation on a hot Summer's day may increase the orienting and appetitive functions of water, which may be accompanied by some relevant relating activity, such as emitting the relational network "Wow, it's so hot today, I really need to find some water." Alternatively, exposure to this relational network (e.g., if another person overheard someone emit this statement) may have a similar impact on the orienting and appetitive functions of water, even though the second individual was not particularly water-deprived. Note, however, that some of the properties of the relational network may differ between the two individuals, because they do not share the same levels of water deprivation. For the water-deprived person, the coherence of the relational network may be relatively high and its flexibility relatively low, but for the non-deprived person coherence may be lower and flexibility higher. Or more informally, it may be difficult to convince the water-deprived individual that they do not need a drink, but relatively easy to persuade the non-deprived person that they are *not* that thirsty. The point we wish to stress here is that the ROE, as a conceptual unit of analysis, appears to facilitate RFT-based analyses of the impact of any contextual variable on the behavior of verbally-able humans with a high degree of precision (i.e., with relatively few scientific terms) in a manner that always stresses the highly dynamical and complex nature of a field of human psychological events.

The verbal self and the ROE. As argued earlier, without AARRing there would be no derived knowing or verbal self, and without derived knowing (or a verbal self) AARRing, at best, would be extremely constrained and limited. Once a verbal self is established in the behavioral repertoire of an individual, it thus becomes an ongoing behavioral event that participates in virtually every ROE. The vast majority of ROEs may be seen as relatively trivial in the grand scheme of things, but the verbal self remains a constant participant in such behavioral events. For example, the relating, orienting, and evoking that occur in the act of switching off a bedroom lamp before going to sleep could be seen as extremely trivial, but it is still a verbal you who turns off the lamp to achieve some outcome (i.e., a good night's sleep). Other ROEs, of course, may be seen as far more fundamental, and are clearly self-focused. For example, the relating, orienting, and evoking that occur in the act of taking an overdose to end one's life could be seen as an attempt to escape, in a very permanent and final way, the very essence of the verbal self. In any case, the constant and iterative daily cycle of ROEing, from the most trivial to the most fundamental of psychological events for a verbal human, could be seen as always involving that very special form of behavior called (derived) knowing.

The concept of the ROE is thus designed to provide a general conceptual unit of analysis, based on RFT, that aims to capture the distinct way in which most humans navigate their psychological worlds. As such, the ROE is based on the RFT view that human "knowing", as a behavior, is only made possible through the evolution of human language and our learning of a specific language through our ongoing interactions with the verbal communities in which we reside from birth through to death. The complexities involved in learning to engage in knowing behavior is thus far from simple and Skinner was indeed perfectly correct, in our view, to refer to such behavior as 'very special.'

Of course, there are many situations in which verbally-able humans fail to report accurately on their own behavior or to identify the causes of their behavior, and it may be tempting to define such behavior as not involving ROEing (because the individual does not *know* what they did or why they did it). According to RFT, however, there is no requirement that AARRing always involves accurately reporting on your own behavior or its causes. Rather, it is the history of learning to report on your own behavior that establishes a verbal self who knows that they know or perhaps do not know something about their own behavior. Or to put it another way, when you report that you do not know if you did something or why you did it, you are reporting that *you know* that you do not know. Thus, the verbal self, or derived knowing, is still at play here. It is also important to understand that accuracy in reporting on one's own behavior may depend on specific properties of that behavior as conceptualized within the HDML framework. For example, it may be that mutually entailed AARRing that is very low in derivation, complexity, and flexibility, and high in coherence, frequently occurs without participating in wider relational networks that are required to report on that mutually entailed AARRing itself (see Barnes-Holmes, Barnes-Holmes, et al., 2017).

More informally, when a behavior becomes extremely well-practiced or highly automatic it may become increasingly difficult to report accurately on that behavior, but you do not cease to be a (verbal) self or become a person who knows nothing (see Hayes, 1984).

The ROE and its implications for process-based explanations of the behavior of verbally-able humans. In proposing the ROE as a new conceptual unit of analysis for RFT, there appear to be important implications for how we use traditional behavioral processes to explain changes in the behavior of verbally-able humans. Consider, for example, the distinction that is sometimes made between reinforcement as an operation and as a process. If a particular pattern of responding produces specific consequences, and the expected outcome is to maintain or increase response rate, then reinforcement *as an operation* has been established. To define reinforcement *as a process* requires, however, that the responding in question is actually maintained or increases as a result of the operation, and not for some other reason. For instance, if response rate increased in the absence of the reinforcement operation (e.g., an extinction burst) then the process of reinforcement has not been observed. The critical point here is that a specific behavioral process is said to occur as the result of a specific operation and not for some other reason. If one accepts the ROE as a conceptual unit of analysis that applies to most if not all behavior produced by verbal humans, reinforcement *alone*, as a behavioral process, cannot be used to explain an increase in response rate for a verbally sophisticated individual. Of course, reinforcement *as an operation* may be applied to the behavior of a verbal human but any resulting increase in response rate cannot be explained simply by appealing to the *process* of reinforcement *per se.* In other words, if we accept the ROE as a (ubiquitous) conceptual unit of analysis (for verbal humans) this requires that we consider the three inseparable properties of the ROE (i.e., relating, orienting, and evoking) in explaining the increase in response rate, and that, by definition, extends beyond the process of reinforcement, *and indeed beyond traditional RFT accounts.*

To appreciate the core argument we are making here, imagine a simple experiment in which a reinforcement contingency is established, for a verbal human, between pressing the space-bar on a computer keyboard and the delivery of points (exchangeable for money). If we observe that response rate increases only when this contingency is operating, a *traditional* RFT explanation might be that the contingency produced a relational networking response (or more informally, a self-generated rule) such as "Pressing the space-bar repeatedly produces lots of points." This relational networking may then be seen as playing a role in increasing response rate in a type of dynamical feedback loop, in which the rule is generated and then *reinforced* by the contingencies. A traditional RFT account suggests that following the rule, rather than space-bar pressing *per se*, was reinforced, and in a sense is broadly consistent with Skinner's (1966) treatise on rule-governed behavior and human problem solving.

In contrast, an *updated* RFT explanation of the increase in space-bar pressing that occurs in our imaginary experiment involves the ROE (relating, orienting, and evoking responses), which of course extends well beyond any simplistic appeal to reinforcement alone. When an RFT analysis involves the ROE, the reinforcement *operation* may still be seen as producing a relational networking response, which then functions as a rule for obtaining points (i.e., the network coordinates with on-going performance), but a relatively complex set of analyses may then follow. For example, an experimental analysis might focus on the four dimensions within the HDML framework. Specifically, as points continue to be earned by following the rule, the network may be seen as increasing in coherence, and reducing in derivation, flexibility, and complexity (see Harte, Barnes-Holmes, Barnes-Holmes, & McEnteggart, 2017, 2018). More informally, as the rule is repeatedly followed (less derived), it may be seen as increasingly accurate or true (coherent), more difficult to change (less flexible), and when stated explicitly it may be simplified to "keep pressing" (less complex). In addition, the concept of the ROE invites analyses that might focus on changes in the orienting and evoking functions of specific features of the experimental context, such as the space-bar, the points feedback on the computer screen, and so on. Such analyses highlight that the behavioral processes involved in a verbally-able human learning to press a space-bar for points certainly extend well beyond the direct reinforcement of space-bar pressing and even beyond the reinforcement of rule-following.

Of course, one could argue that the ROE as a unit of analysis is not necessary, in that the space-bar pressing in the foregoing example could be readily explained simply by appealing to the process of reinforcement. Although this type of (simple) analysis may be sufficient for certain purposes, it remains the case that the research outlined earlier, involving the IRAP and the DAARRE model, call for ROE-based analyses or at least analyses that grapple with the types of behavioral properties highlighted by the ROE. Or in other words, we do need to consider the cluster of variables specified by the ROE, and their dynamic interplay, if we wish to predict-and-influence the behavior of verbal humans when they respond in accordance with relatively simple relational networks. Furthermore, ROE-based analyses seem to be required even when verbal humans respond in very brief periods of time, as is the case with the IRAP.

If one accepts the ROE as a conceptual unit of analysis that may be applied to most if not all instances of behavior produced by verbal humans, there are immediate implications for dealing with behavioral processes in the context of psychotherapy. In broad terms, the core implication is that the assessment and treatment of human psychological suffering will always be focused on AARRing as conceptualized through the ROE, or some version thereof. Or to put it another way, psychotherapy (for verbal humans) will always be focused on that very special form of behavior called knowing. Part 2 of the current article provides clinically relevant examples that speak directly to this implication.

Part 2

The HDML Framework and a Process-Based Approach to Human Psychological Suffering

Behavioral science has always concerned itself with the processes of learning, adaptation, and behavioral change. Until the late 1960s or early 1970s, a widely-held assumption was that these behavioral processes, broadly speaking, were common to both nonhumans and humans. This assumption was reflected in the earliest translational research associated with behavioral psychology. The famous study by Watson and Rayner (1920) in which they created and “treated” a phobia in a young child, using the processes of classical conditioning and extinction, which had been identified and studied by Pavlov using dogs (1897, 1902), provides a clear-cut example. Other examples, of course, abound in the literature, including the study of learned helplessness (Seligman, 1974), inhibition (Wolpe, 1968), and fear generalization (Lashley & Wade, 1946), each of which has been used in experimental analogs of both human and nonhuman ‘psychopathology’. The continuity assumption, at the level of psychological processes, from animals to humans has not been without value, but it remains that – an assumption, not an empirical fact.

While many scientists assume that there are differences between human and nonhuman psychological processes (e.g., Chomsky, 1959; Pinker, 1994; Premack, 2007; Sidman, 1994; Wilson, Hayes, Biglan, & Embry, 2014), there remains highly-regarded cutting-edge process-oriented clinical research that fails to grapple meaningfully with these differences. For example, recent work by Craske and colleagues on an inhibitory learning approach to maximizing the impact of exposure therapy (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014) draws heavily on basic research conducted with nonhumans (e.g., Bouton, 1993). The underlying assumption thus appears to be that psychotherapy should be based on, and needs to target, inhibitory learning processes that are common to both human and nonhuman species. In pointing to the work of Craske, we are not questioning its quality or effectiveness, and indeed we applaud Craske's focus on processes in developing therapeutic interventions. But, we believe that a more complete process-based approach to *human* psychological suffering and its treatment should be informed by the empirical and conceptual analyses covered in Part 1 of the current article. We will now present some relevant examples of how the HDML framework could be of value in conceptualizing the clinical assessment and treatment of human psychological suffering.

Before continuing, it is important to understand that the HDML framework (and the ROE as a core unit of analysis) should not be viewed as a type of 'protocol' or manual to be followed. Our aim here is simply to illustrate how thinking in terms of the HDML framework provides the foundation for conducting what we have referred to previously as *verbal functional analyses* (Barnes-Holmes, Boorman et al., 2018). Thus, what we offer here is not entirely new, but constitutes the next step in our on-going attempt to connect the basic science with the application and practice of that science through the development of well-defined behavioral processes. We will attempt to do this by working through two brief case summaries, which involved a number of the current authors directly. Across these sections, we illustrate how the HDML framework facilitates a functional-analytic approach to assessment and treatment that is rooted in a behavioral process-oriented therapy.

The HDML Framework and its Implications for Behavioral Process-based Therapy

Case Example 1. Consider the case of Sarah, a married woman in her 30s with two young children. Sarah is one of three female children, all of whom were victims of familial sexual trauma. Sarah presents with depression and initially talks a great deal about a feeling in the pit of her stomach typically noticed first thing in the morning, which is always taken as a signal that she will be unable to fully meet the demands of that day. That is, Sarah reveals that this sensation in her stomach is associated with general feelings of inadequacy. She also focuses strongly on the fact that because her feelings of inadequacy straddle the most important areas of her life, parenting, being a wife, and being a professional, she must be genuinely inadequate as a human being. She adds that she has felt completely inadequate in these areas for many years, including, for example, the belief that she should not have had children. She reports that she had seen many therapists, had tried self-help techniques, and had been on a regime of medication for many years, but these made no difference to her intense feelings of inadequacy.

Within the framework of the HDML, and through the lens of the ROE more specifically, we could conceptualize some of the therapeutic interactions involving Sarah as follows. First, the feeling in the pit of her stomach has strong orienting functions (i.e., the feeling is highly salient). This feeling participates in relational networking involving the deictic-I (verbal self), guilt, and inadequacy. The orienting toward the stomach sensation combined with this relational networking has aversive evoking properties. More informally, when Sarah wakes up with that feeling in the pit of her stomach, she quickly interprets it as a reason to feel inadequate, which of course is aversive to her.

In terms of the dimensions of the HDML framework, Sarah's belief in her own inadequacy could be described as highly coherent because she reports that it spans all three of the most important areas of her life. Complexity may also be deemed relatively high because there is a rich and extended narrative that relates her widespread feelings of inadequacy back to her history of childhood abuse, including all of the attempts she has made to struggle with her psychological suffering. Derivation could be described as relatively low because the complex narrative around her inadequacy had been well rehearsed over many years. Flexibility could also be described as relatively low in that Sarah reported that none of her attempts over the years to reduce her feelings of inadequacy had any effect.

During the course of therapy, the therapist explored the orienting functions of the stomach tightening and, in particular, the extent to which they participate in the relational networks of 'self', 'inadequacy', and 'failure'. For example, Sarah was asked to reflect upon the extent to which this specific somatic feeling upon waking up tended to dominate every aspect of her life every day. In doing so, the therapist aimed to reveal how central this struggle had become in her life. As such, within the HDML framework, the stomach tightening and feelings of inadequacy could be deemed to have appetitive evoking functions.[10] The aim in therapy was thus to undermine the appetitiveness of these particular networks. This was achieved by working through all of the ways orienting toward these networks had not worked for Sarah (i.e., failed to produce any improvement in her wellbeing).

During this work, the therapist was careful not to establish aversive evoking functions for the relational networks of the verbal self (or deictic-I). In other words, the therapeutic work sought to avoid creating self-recrimination and/or self-loathing for Sarah. Specifically, the therapist sought to establish the functions of 'hopelessness' or 'unworkability' in the relational networks of inadequacy. This provided the context for alliance-building by coordinating Sarah's legitimate attempt to escape from the orienting/evoking functions of the relational networks with the therapist's deictic-I, with the use of phrases such as "that's what I would have done too" and "I can see how that makes complete sense."

In exploring the relationship between the verbal self and the relational networks of inadequacy and failure etc., the term "disgust" and its coordination with the verbal self emerged. It was at this point that the therapist felt that a functionally aversive relational network had been identified, and thus therapy began to focus on disgust rather than inadequacy. This involved highly sensitive therapeutic work in which therapist and client revisited childhood abuse episodes together, exploring verbally the extent to which she felt (erroneously) that she was responsible for what happened to her. In other words, the therapist

---

[10] In suggesting that the relational networks involving stomach tightening and feelings of inadequacy etc. are appetitive, we mean this in a purely functional-analytic way. That is, Sarah is quite willing to use these terms to describe her struggle, but as we shall see subsequently there are other terms that she is far less willing to use in this context, which would be more appropriately described as aversive.

focused on elaborating the networks around disgust and responsibility, with unconditional support and with the therapist showing no sense of pulling back. In so doing, the therapist was aiming to transform the orienting and evoking functions of the ‘disgust’ network, so that Sarah became increasingly willing to engage with these previously highly-avoided issues of disgust and perceived responsibility around her childhood trauma. For example, the therapist aimed to decrease coherence in the relational networking between responsibility and the verbal self, thus allowing new derivations to emerge, such as ‘If I was not responsible, I am not disgusting’. In time, this provided her with opportunities to engage with her life outside therapy without the narrow and inflexible repertoires that were being evoked by the networks of self-blame and disgust.

Case Example 2. Consider the case of Lisa, a married woman in her 40s. Lisa presents with anxiety and a specific phobia of contamination from people and inanimate objects (e.g., seats on public transport). Lisa is overwhelmed by this fear and related intrusive thoughts, and their manifestation almost everywhere in public, leading to panic attacks and the need to leave public places to return home immediately. During the course of therapy, it becomes apparent that Lisa focuses strongly on the possibility that if she is contaminated, she could potentially contaminate other people such as her children, and present a danger to them. This situation has created problems in her intimate relationships and rendered her unable to attend any events that require use of public facilities (e.g., saunas, swimming pools, and even shopping centers). She reports that she has tried various self-help techniques and taken medication for anxiety for a number of years, but reported that both were largely ineffective.

Within the framework of the HDML, and the ROE more specifically, we could conceptualize some of the therapeutic interactions involving Lisa as follows. First, there were very strong orienting functions for any stimulus that was potentially contaminated (e.g., a restaurant table that had not been wiped clean). Once identified as such, that stimulus would have highly aversive functions, which would then spiral her into highly complex relational networking around the danger presented and should she become contaminated the danger she would potentially pose to others, especially her children. In contexts from which she could not escape, she would engage in further relational networking that sought to alleviate her anxiety with reasons why she would not likely become contaminated (e.g., asking a waiter to move her to a different table).

In terms of the dimensions of the HDML framework, Lisa's fear of contamination could be seen as involving relational networks that are very high in coherence, because the fear is described as overwhelming, highly intrusive, and induced by stimuli that are found everywhere in her environment. Complexity at the beginning of therapy may have been defined as relatively low because the relational networking around her anxiety was linked to the orienting functions of potentially contaminating stimuli in the environment and little else. As we shall see, however, during the course of therapy, the networking around her psychological struggle appeared to increase in complexity, as the narrative around contamination was elaborated. Derivation could be described as low because Lisa described her struggle as going back years. Flexibility could also be described as low in that Lisa indicated that her use of self-help techniques and medication were ineffective.

Early in therapy, the therapist explored the orienting and evoking functions around contamination, particularly the aversiveness of the evoking functions, and the extent to which this aversiveness participated in networking around the need to act (e.g., going home immediately). In doing so, a highly inflexible relation between the networks around contamination, and the need to act, emerged. Specifically, the deictic-I network acquired the functions of being contaminated, and also being dangerous, to the extent that she could contaminate her children.

At this point, an effort was made in therapy to create a relatively coherent relational network or narrative pertaining to Lisa's contamination fears that the therapist could use to undermine the behavioral control functions of the contamination network per se. Specifically, the orienting, evoking, and relating involved in contamination were used to construct a complex relational network for Lisa. Specifically, this network emphasized a sequence of causal relations, in which 'noticing' a potentially contaminating stimulus immediately evoked thoughts of being contaminated/dangerous, followed by highly aversive feelings of being overwhelmed, which typically led to Lisa seeking to return home immediately. This sequential pattern of relational responding was rendered by the therapist into a hierarchical network by labeling it as a "stream" of reactions and behaviors that Lisa engaged in.

Once the functional properties of the word "stream" had been established as a way to talk about her psychological struggle, the therapist then validated the stream by highlighting the coherence between the stream and how the therapist herself would react if she was in a similar stream. Engaging in this specific aspect of alliance-building allowed the therapist to help Lisa to explore and elaborate the functions of the stream, by considering how it may function as self-harm rather than as self-protection. At this point, the therapist used the stream network to create a metaphor that sought to undermine the behavioral control functions of the stream itself. Specifically, the stream was contrasted with a rock upon which Lisa could sit and observe the stream, thereby distancing herself from the behavioral control functions of the struggle. For example, the therapist suggested that Lisa could choose to sit on the rock and observe the stream, but if she then chose even to dip her toe into the stream, there was a danger that she would be tempted to jump in, and very quickly find herself engaging in what looked like 'self-protective' behavior to prevent herself from becoming overwhelmed.

Having reached this stage of therapy, the aim was to increase coherence in the new way of talking about and responding to the struggle, and to reduce derivation and flexibility, so that the original behavioral control functions of the stream continued to be undermined in a variety of contexts. One of the key ways in which this was done was by asking Lisa to record herself with her phone every time she saw a potentially contaminated stimulus. During the next session, she and the therapist would then watch the recording together and explore Lisa's reactions to herself on the recording. This exercise provided multiple exemplars that served to increase coherence and reduce derivation and flexibility in the newly-established pattern of relational responding pertaining to the reduced behavioral control functions of the struggle itself. Another way in which the therapist sought to increase coherence and decrease derivation and flexibility was to openly share this new narrative about her struggle with other people in her life. In effect, these exercises allowed Lisa to practice this new way of relating to her struggle in a wide variety of contexts.

Summary of general strategy for an updated RFT approach to behavioral process-based therapy. Having described two case summaries that provide examples of how an updated version of RFT could approach behavioral process-based therapy, it would seem useful to broadly outline the general approach we have taken here. In doing so, we should emphasize that this is very much a work in progress and, as such, reflects our thinking at the time of writing. We thus anticipate that what we offer here, both conceptually and clinically, will continue to develop and be refined over the coming years. Consistent with this view, the following summary of the general strategy will be presented below in bullet point form, so that, in closing, the reader (and particularly a clinician) has a relatively concise series of take-home messages with regard to RFT-driven process-based psychotherapy. In presenting the summary in bullet point form, it is important to understand that we are not arguing for a fixed linear step-by-step sequential approach to therapy. Rather, what we offer are key areas to focus on during the course of therapy.

- Early in therapy (and throughout if necessary), the therapist tends to focus on the orienting functions, particularly those that have high levels of behavioral control (e.g., Lisa's orienting toward potentially contaminating stimuli and the need to avoid them). These functions can arise from internal or external sources that draw the client's attention, including: somatic (e.g., palpitations); perceptual (e.g., hearing voices); complex relational responses (e.g., ruminating)[11]; and external events (e.g., environmental triggers). An important part of therapy involves helping the client to identify the relevant events that possess the critical orienting functions that both the therapist and client will focus on throughout the therapeutic process.
- In focusing on orienting functions as suggested above, we are not denying that evoking functions will also be involved. For example, when a client wakes up orienting towards a strong somatic sensation (e.g., Sarah's feeling in her stomach), the sensation itself will likely have strong (typically aversive) evoking functions. Hence, it is difficult for the therapist to engage the client in relating with regard to orienting, without also engaging evoking functions. For example, if the same client is asked "What do you tell yourself about that sensation?", they might answer "I just can't bear it" (indicating strong aversive evoking functions). Critically, what the therapist is asking the client to do here is to engage in relational responding with regard to both orienting and evoking functions of the specific sensation. More informally, the therapist is seeking to get a sense of the story (i.e., complex relational networking) that is built around the sensation and its evoking properties. Noting this general strategy serves to highlight why the ROE was presented earlier as a dynamical system, consisting of three inseparable properties.

[11] The reader should note that orienting functions may arise from complex relational responses, which is why the ROE is presented earlier as a non-linear, dynamical system that is constantly in motion as human beings navigate their internal and external psychological worlds.

- An important purpose of exploring the relational networking around orienting and evoking functions is to identify the extent to which such networking may be seen as relatively coherent versus incoherent. For example, a client might say that when they have a specific sensation, they “just feel like they want the world to end”, but soon thereafter the same client might say “but I know it’s not so bad really.” Identifying a clear example of relational incoherence such as this may indicate that the apparently aversive evoking functions of the sensation also have appetitive evoking functions. That is, orienting so strongly towards a particular sensation, and its aversive functions, but at the same time relating less aversively or even appetitively toward the sensation is a pattern (of relational incoherence) we actively search for in therapy. In elaborating the networking around these functions in therapy, a key source of psychological suffering that the client may have avoided for many years can be identified (e.g., elaborating Sarah’s ‘inadequacy’ network such that it relates to the ‘self-blame’ network). It is important to note than an aim of therapy is *not* to create a coherent narrative for the client, but to use incoherence as a flag for a likely source of distress.
- Early in therapy (and again later if necessary), it seems wise to focus on the appetitive (or at least less aversive) evoking functions, with a view to facilitating therapeutic interactions. Through such engagement, it becomes increasingly probable that the client will engage more readily with the aversive evoking functions.
- Once the client is engaging with aversive evoking functions, it becomes possible to elaborate the networking around the relationships between the appetitive *and* aversive evoking functions. Doing so helps to establish networking around the deictic-I as containing both appetitive and aversive evoking functions. More informally, this new and increasingly coherent narrative helps the client to both *observe and own* the

distress-inducing contradictions that have existed in their own behavior[12]. Critically, this therapeutic process serves to undermine the behavioral control functions that these contradictions previously had for them in their lives. In short, therapy is focused on freeing the client from being a hostage to their own history.

In providing the foregoing series of take-home messages, the single generic point that we would emphasize is that an RFT-informed version of process-based therapy requires a focus on complex relational networking rather than on *merely* framing. The emergence of the MDML framework, the DAARRE model, and their integration in the HDML framework, which generated the ROE as a conceptual unit of analysis, has facilitated this new focus. As noted earlier, the original Kantorian flavor to RFT re-emerges when the analysis of AARR involves studying a field of behavioral (verbal) interactants. In our view, this renewed emphasis on the field-based nature of RFT has greatly enhanced our ability to develop the beginnings of a highly process-oriented therapy, such as we have described in the current article.

Concluding Comments

Having been dominated for decades by the medical model of human psychological suffering and a syndromal approach to identifying particular classes of that suffering, we are now entering a more process-oriented era. Naturally, this shift in emphasis sits well with the behavior-analytic tradition, in that behaviors are not seen to be proxies for underlying pathologies or syndromes. In embracing a more process-oriented approach to human psychological suffering, behavioral science will need to step up to the monumental challenge of dealing with the complexities involved in such suffering. In so far as those processes are always at least partly verbal, the challenge requires that behavioral science continues to

[12] We suspect that establishing a new, coherent narrative around the deictic-I is a key aim in many therapeutic regimes.

construct an account of human language and cognition, and connects that account to therapeutic assessment and treatment. The enormity of the challenge that confronts clinical psychology, and psychological science in general, cannot in our view be underestimated. The challenge is not simply one of attending more carefully or systematically to replication, reproducibility, or measurement (Hayes et al., 2019). It is not simply a matter of embracing new technologies or increasingly sophisticated methodologies and analytic methods, quantitative and/or qualitative (e.g., Maxwell, Lau, & Howard, 2015). Also, the challenge will not be met simply by encouraging basic and applied researchers to work more closely with practitioners (e.g., Blount, Bunke, & Zaff, 2000). And the problem psychology faces will not be addressed by attempting to align itself more closely with other sciences (e.g., Schwartz, Stapp, & Beauregard, 2005). Each of these areas of concern are of course important in their own right and deserve our attention. In our view, however, the fundamental challenge is one that calls for a radical and genuinely paradigm-shifting perspective on what the science of psychology should consider its fundamental subject matter, and how that subject matter is conceptualized and studied.

The paradigm shift we propose involves removing the person or individual as an active agent from the analysis of psychological events, leaving nothing but complex clusters of behavioral functions, defined in terms of historical and current contextual variables. The DAARRE model and its integration into the MDML framework provides, we believe, an example of this general strategy. There is no person or individual in the DAARRE model, the HDML framework, or the ROE unit more generally. Instead, the psychological event(s) is presented as a field of verbal interactants. The individual elements do not exist independently of the other interactants; rather they are actualized by their participation in a field of interactants. In Figure 1, for example, the + orienting function for the label stimulus “Color” is defined, in part, relative to the - orienting function for the label stimulus “Shape.” The field of interactants that are actualized in the analysis of a specific IRAP performance thus provide the very definition of a psychological event and the psychological event is the field – they are one and the same "thing." There is no person, as such, *contained* within the field; in a sense the person *is* a constantly changing or actualizing field of verbal interactants.

From this perspective, a clinician is not "assessing" or "treating" a "sick," "broken," "delusional," or "depressed" individual. Rather the therapy is conceived of as a complex and dynamical field of verbal interactants as the therapist and client engage inside and outside of the therapy room. Of course, the critically important question then becomes, does this way of way of conceptualizing clinically relevant psychological events lead to improvements in behavioral prediction-and-influence, with precision, scope, and depth, relative to alternative ways of talking about psychological suffering? The answer to this question will take years if not decades to answer, but the general strategy of searching for a single overarching conceptual framework for analyzing psychological events in general, and psychological suffering in particular, seems like an investment worth making.

Compliance with Ethical Standards

Funding: This research was conducted with funding from the Odysseus Programme at FWO, awarded to the first author.

Conflict of Interest: The authors declare that there is no conflict of interest.

Ethical approval: All procedures in the current study were in accordance with the ethical standards of the institutional research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants.

Availability of Data and Materials. The current article is conceptual and therefore has no associated data or materials.

References

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*(6), 925.

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., & Holmes, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record*, *41*(1), 19-31.

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*(2), 251-265.

Barnes-Holmes, D., Barnes-Holmes, Y., Finn, M., Harte, C., Kavanagh, D., … , Quak, M. (under review). The dynamics of arbitrarily applicable relational responding and an updated version of relational frame theory. *Perspectives on Behavioral Science.*

Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., & McEnteggart, C. (2017). From the 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,* 434-445.

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(4), 497-516.

Barnes-Holmes, D., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (2018). Derived stimulus relations and their role in a behavior-analytic account of human language and cognition. *Perspectives on Behavior Science, 41*, 155-173.

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*(1), 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*. New York (pp. 103-117), New York, NY: Plenum.

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

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(3),* 531-558.

Barnes-Holmes, Y., Boorman, J., Oliver, J. E., Thompson, M., McEnteggart, C., & Coulter, C. (2018). Using conceptual developments in RFT to direct case formulation and clinical intervention: Two case summaries. *Journal of Contextual Behavioral Science, 7,* 89-96.

Blount, R. L., Bunke, V. L., & Zaff, J. F. (2000). The integration of basic research, treatment research, and clinical practice in pediatric psychology. In D. Drotar (Ed.), *Handbook* *of research in pediatric and clinical child psychology*. Issues in Clinical Child Psychology. Boston: Springer.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80.

Chomsky, N. (1959). A review of BF Skinner's verbal behavior. *Language*, *35*(1), 26-58.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10-23.

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*(2), 179-197.

Dymond, S., & Barnes, D. (1994). A transfer of self- discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behavior*, *62*(2), 251-267.

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*(2), 163-184.

Dymond, S., & Barnes, D. (1997). Behavior-analytic approaches to self-awareness. *The Psychological Record*, *47*(2), 181-200.

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*(2), 309-321.

Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the single-trial-type-dominance-effect on the IRAP: Developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *The Psychological Record*, *68*, 11-25.

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.

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.*

Hayes, S. C. (1984). Making sense of spirituality. *Behaviorism,* 99-110.

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

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.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst*, *9*(2), 175-190.

Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy, 117,* 40-53.

Hayes, S. C., & Sanford, B. T. (2014). Cooperation came first: Evolution and human cognition. *Journal of the Experimental Analysis of Behavior*, *101*(1), 112-129.

Hayes, S. C., Sanford, B. T., & Chin, F. T. (2017). Carrying the baton: Evolution science and a contextual behavioral analysis of language and cognition. *Journal of Contextual Behavioral Science*, *6*(3), 314-328.

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science*, *7*(1), 37-50.

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*(2), 207-227.

Hughes, S., & Barnes-Holmes, D. (2014). Associative concept learning, stimulus equivalence, and relational frame theory: Working out the similarities and differences between human and non-human behavior. *Journal of the Experimental Analysis of Behavior*, *101*(1), 156-160.

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, UK: 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, UK: Wiley-Blackwell.

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), 1-15.

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*(3), 395-403.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70* (6), 487-498.

Lashley, K. S., & Wade, M. (1946). The Pavlovian theory of generalization. *Psychological Review*, *53*(2), 72.

Lattal, K. A. (1975). Reinforcement contingencies as discriminative stimuli 1. *Journal of the Experimental Analysis of Behavior*, *23*(2), 241-246.

Leech, A., Barnes-Holmes, D., & Madden, L. (2016). The implicit relational assessment procedure (IRAP) as a measure of spider fear, avoidance, and approach. *The Psychological Record*, *66*(3), 337-349.

Leech, A., Barnes-Holmes, D., & McEnteggart, C. (2017). Spider fear and avoidance: a preliminary study of the impact of two verbal rehearsal tasks on a behavior–behavior relation and its implications for an experimental analysis of defusion. *The Psychological Record*, *67*(3), 387-398.

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*(2), 201-239.

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*(3), 349-365.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141-144.

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *The Psychological Record*, *54*(1), 115-144.

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*(3), 437-460.

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.

Pavlov, I. P., & Thompson, W. H. (1897/1902). *The work of the digestive glands*. Charles Griffin.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377-410.

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*(3), 275-301.

Ruiz, F. J., & Luciano, C. (2011). Cross- domain analogies as relating derived relations among two separate relational networks. *Journal of the Experimental Analysis of Behavior*, *95*(3), 369-385.

Schwartz, J. M., Stapp, H. P., & Beauregard, M. (2005). Quantum physics in neuroscience and psychology: A neurophysical model of mind-brain interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1309-1327.

Seligman, M. E. (1974). *Depression and learned helplessness*. John Wiley & Sons.

Shimp, C. P. (1982). On metaknowledge in the pigeon: An organism's knowledge about its own behavior. *Animal Learning & Behavior*, *10*(3), 358-364.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research*, *14*(1), 5-13.

Sidman, M. (1994). *Stimulus equivalence: A research story*. Boston, MA: 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*(1), 5-22.

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.

Skinner, B. F. (1974). *Walden two*. Hackett Publishing.

Stewart, I., & Barnes-Holmes, D. (2004). Relational frame theory and analogical reasoning: Empirical investigations. *International Journal of Psychology and Psychological Therapy, 4(2),* 241-262.

Villatte, M., Villatte, J. L., & Hayes, S. C. (2015). *Mastering the clinical conversation: Language as intervention.* New York, NY: The Guilford Press.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1.

Wilson, D. S. (2012). Making contextual behavioral science part of the united ivory archipelago. *Journal of Contextual Behavioral Science*, *1*, 39-42.

Wilson, D. S., Hayes, S. C., Biglan, A., & Embry, D. D. (2014). Evolving the future: Toward a science of intentional change. *Behavioral and Brain Sciences*, *37*(4), 395-416.

Wolpe, J. (1968). Psychotherapy by reciprocal inhibition. *Conditional Reflex: A Pavlovian Journal of Research & Therapy*, *3*(4), 234-240.

Table 1

A Multi-Dimensional Multi-Level (MDML) Framework Consisting of 20 Intersections Between the Dimensions and Levels of Arbitrarily Applicable Relational Responding.

LEVELS	**DIMENSIONS**			
	Coherence	**Complexity**	**Derivation**	**Flexibility**
Mutually Entailing	Analytic Unit 1	Analytic Unit 2	…	…
Relational Framing	…	…	…	…
Relational Networking	…	…	…	…
Relating Relations	…	…	…	…
Relating Relational Networks	…	…	…	Analytic Unit 20

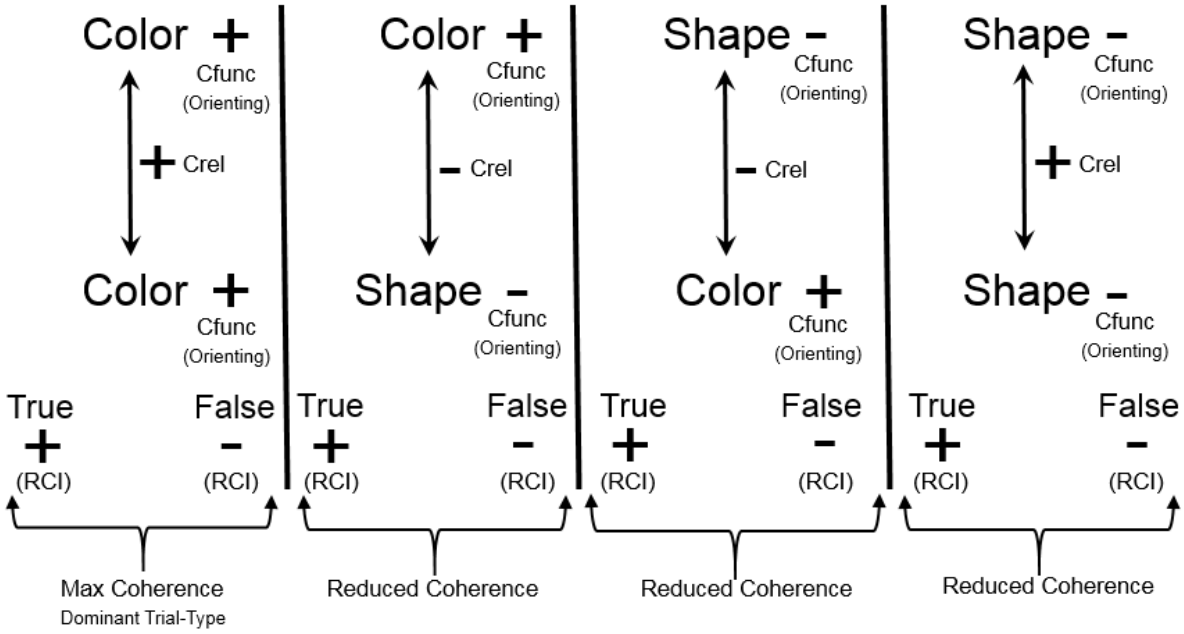

Figure 1. The DAARRE model as it applies to the “Shapes-and-Colors” IRAP. The positive and negative symbols refer to the relative positivity of the transformation-of-function property (Cfunc), for each label and target, the relative positivity of the entailment property (Crel), and the relative positivity of the relational coherence indicator (RCI) in the context of the other Cfuncs, Crels, and RCIs.

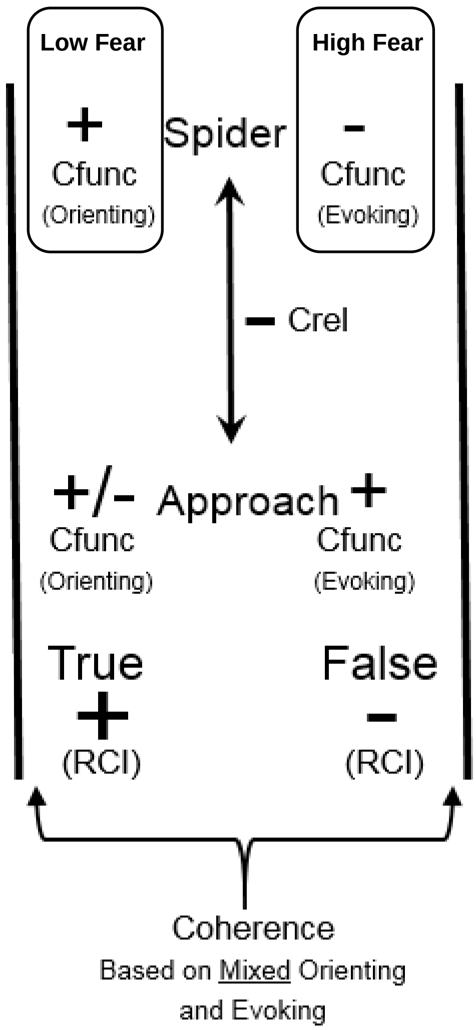

Figure 2. The DAARRE model as it applies to the Spider-Approach trial-type of the "Pets-and-Spiders" IRAP. The terms "Low Fear" and "High Fear" indicate the Cfuncs that are likely to dominate for individuals who are low (orienting) versus high (evoking) in spider fear. The assumption that the orienting functions of approach relative to avoidance descriptors would not differ dramatically in the context of this particular IRAP is indicated by the symbol "+/-".

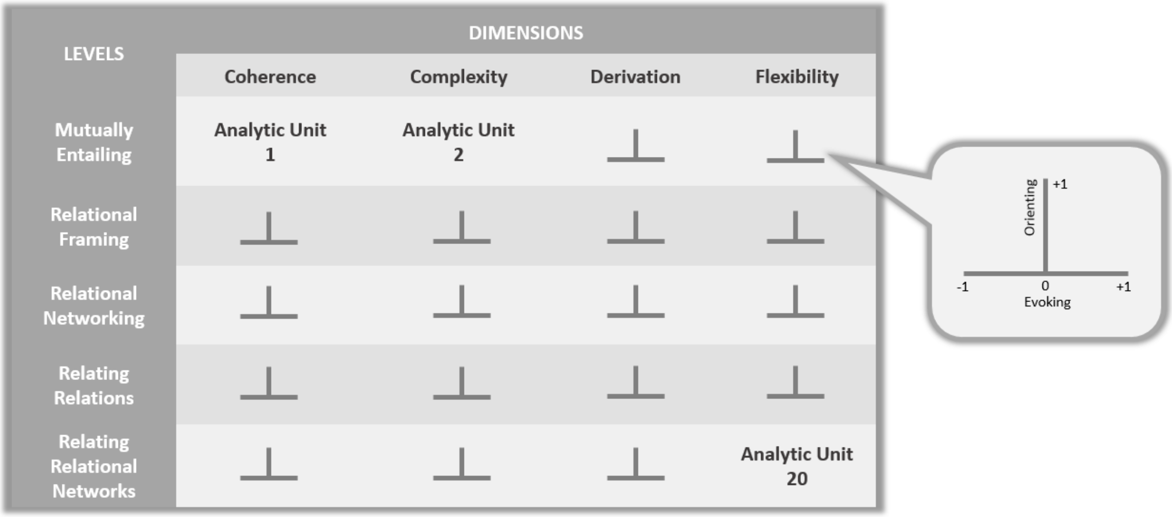

Figure 3. A **hyper**-dimensional multi-level (HDML) framework consisting of 20 intersections between the dimensions and levels of arbitrarily applicable relational responding, combined with the dimensions of orienting and evoking from the DAARRE model.