Back to the Future with an Up-dated Version of RFT: More Field than Frame?

Dermot Barnes-Holmes[1,2] Yvonne Barnes-Holmes[1], Ciara McEnteggart[1], & Colin Harte[1]

[1]Department of Experimental, Clinical, and Health Psychology, Ghent University

[2]School of Psychology, University of Ulster

Corresponding Author:

Dermot Barnes-Holmes

Department of Experimental, Clinical, and Health Psychology

Ghent University

Henri Dunantlaan, 2

B-9000 Ghent

Belgium

Email: Dermot.Barnes-Holmes@ugent.be

Authors' Note This chapter was prepared with the support of an Odysseus Group 1 grant awarded to the first author by the Flanders Science Foundation (FWO). Some of the material presented in the current chapter appears in a recent article published in *The Psychological Record*, entitled, "Up-dating RFT (more field than frame) and its implications for process-based therapy" (Barnes-Holmes, Barnes-Holmes, & McEnteggart, 2020); Copyright permission granted. Correspondence concerning the current chapter should be sent to Dermot.Barnes-Holmes@ugent.be.

Abstract

The current chapter presents an overview of a line of research that focuses on the behavioral dynamics of arbitrarily applicable relational responding (AARRing), and the implications of this research for the on-going development of relational frame theory (RFT) itself. Specifically, the integration of two recent conceptual developments within RFT are described. The first of these is the multi-dimensional, multi-level (MDML) framework and the second is the differential arbitrarily applicable relational responding effects (DAARRE) model. Integrating the MDML framework and the DAARRE model emphasizes the transformation of functions within the MDML, thus yielding a *hyper*-dimensional, multi-level (HDML) framework for analyzing the behavioral dynamics of AARRing. The HDML generates a new conceptual unit of analysis for RFT in which *relating*, *orienting,* and *evoking* (ROEing) are seen as involved in virtually all psychological events for verbally-able humans. Some of the implications of the ROE as a unit of analysis for RFT are explored, including the idea that it may be useful to conceptualize the dynamics of AARRing as involving a field of verbal interactants.

Key words Dynamics, RFT, MDML, DAARRE, HDML, relating, orienting, evoking, field of interactants

In writing a concluding chapter for the current volume on RFT it would be tempting simply to provide a summary or overview of the most recent RFT research. Given that at least some of that research will have been covered in earlier chapters, it seemed more appropriate to focus on a particular line of research that has involved considerable conceptual development in the theory itself. Specifically, we have argued that the metaphor of the frame might be usefully replaced by the metaphor of the "field" (of verbal interactants). In making such an argument we are not proposing an alternative theory, or even calling for RFT to be renamed relational *field* theory. In fact, if anything, we are harking back to the earliest days in the development of RFT, when it was focused more on the analysis of complex relational *networks* involved in rule-governed behavior (see Hayes & Hayes, 1989), rather than as a theory of equivalence relations and the analysis of individual frames (see Hayes, 1991). Indeed, 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), and in this sense we are suggesting that the future of RFT appears to involve, paradoxically, going back in time to its conceptual roots.[1]

We should be clear, at this point, that our view of RFT as "going back to the future" (i.e., with an emphasis on the interbehavioral "field") arose during the course of a relatively intense period of empirical research; and in the conceptual and philosophical analyses that accompanied that work, both in day-to-day discussions that took place in our research group and in the writing of various articles and book chapters throughout that period. We will not, therefore, present the case for a more field-like emphasis at the beginning of the current chapter, but will present it towards the end when the relevant empirical and conceptual

[1] The term "relational frame" appears in an article in *The Behavior Analyst* by Brownstein and Shull (1985), and is used in a discussion of "rule-governed behavior." However, the use of the concept of relational frame is treated as a complex type of discriminative stimulus and no reference is made to the literature on equivalence relations. In this sense, it appears that Brownstein and Shull were using the term relational frame as a composite of the terms "autoclitic frame" and "relational autoclitic" as they are used in the book *Verbal Behavior* (Skinner, 1957), which of course pre-dates the study of equivalence or derived relations.

developments have been outlined. In this way, we hope to convey to the reader a sense of the intellectual journey that brought us to the conclusion that a more field-like emphasis could be extremely useful in taking RFT forward as a modern behavior-analytic account of human language and cognition.

Background to the Current Chapter

The early study of equivalence class formation, and derived relational responding more generally, tended to focus on "demonstration-of-principle" research. That is, studies often aimed to produce a particular pattern of derived relational responding that was either present or absent during a block of test trials. Of course, such a focus was necessary in the early stages of the research program, but we have been arguing more recently for the development of concepts and methods for exploring the behavioral dynamics of derived relational responding itself (e.g., Barnes-Holmes, Barnes-Holmes, Luciano, & McEnteggart, 2017; Barnes-Holmes, Finn, Barnes-Holmes, & McEnteggart, 2018). To this end, the current chapter will cover our attempt to integrate two recent conceptual developments within RFT. The first of these is the multi-dimensional, multi-level (MDML) framework (see Barnes-Holmes, 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, Barnes-Holmes, & McEnteggart, 2018).

The integration of the MDML framework and the DAARRE model appears to capture the two core defining properties of AARRing itself, entailment relations and the derived transformation of functions (Barnes-Holmes, Barnes-Holmes, & McEnteggart, 2020).[2] These two properties were contained in both the MDML framework and the DAARRE model, as

[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 AARRing in general, not relational frames specifically. In the context of the MDML, therefore, it seems wise to refer to two general properties of AARRing, entailment and transformation functions.

originally proposed, but particularly in the case of the MDML framework the emphasis appeared to be on entailment alone. The integration thus attempts to place a greater emphasis on function transformation within the MDML framework by drawing on the DAARRE model, thus yielding what is now called a *hyper*-dimensional, multi-level (HDML) framework for analyzing the dynamics of AARRing. The HDML then led us to propose a new conceptual unit of analysis for RFT in which *relating*, *orienting,* and *evoking* are seen as being involved in virtually all psychological events for verbally-able humans; this new analytic is referred to as the ROE (pronounced "row") and will also be covered in the current chapter (Barnes-Holmes, et al., 2020).

The MDML Framework

In an effort to systematize the RFT account and emphasize the relevant behavioral dynamics, the MDML was offered 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) mutually entailing, to (ii) simple networks involved in framing, 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 should not be seen as 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 grape is larger than an orange" 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).[3] *Derivation* refers to how well practiced a particular instance of AARRing has become. Specifically, when a pattern of AARRing is derived for the first time it is, by definition, highly derived (i.e., novel or emergent), and thus derivation reduces as that pattern becomes more practiced (i.e., less and less novel or emergent). *Flexibility* refers to the extent to which a given instance of AARRing may be modified by current contextual variables. Imagine a young child who is asked to respond with the wrong answer to the question "Which is bigger, a grape or an orange?" The more readily this is achieved, the more flexible the AARRing.

[3] Relational complexity (and indeed the other dimensions) may be defined along more than one dimension, such as number of a 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).

The MDML framework 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 broader than B, then B will be narrower 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 narrower than A in the absence of programed 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

The Basic Model

As noted at the beginning of the current chapter, the MDML framework appears to be very much focused on the entailment relations (or Crel properties) of AARRing. The function-transformation effects (or Cfunc properties) of AARRing have always been assumed within the MDML framework, because RFT defines AARRing itself in terms of both properties. Nevertheless, it seemed important to incorporate function-transformation into the MDML framework in a relatively explicit manner. This objective was 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 between the DAARRE model and 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" (or STTDE) 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 chapter; but see Finn, et al. 2018; see also, Kavanagh, Barnes-Holmes, Barnes-Holmes, McEnteggart, & Finn, 2018).

In attempting to explain the STTDE 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 itself 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 (see Pinto, de Almeida, & Bortoloti, 2020, for a recent study that provides evidence for differential orienting responses on the IRAP using eye tracking as a measure).

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.

Insert Figure 1 Here

As noted above, the STTDE 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 sign) 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 STTDE (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 STTDE.

At this point, it is critically important to understand 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. Thus, for example, the maximal coherence of the *Color-Color* trial-type is based on what we assume are the likely behavioral histories prevailing within the relevant language community. Consequently, it may well be possible to establish a STTDE for one of the other trial-types. Imagine, for example, that the same IRAP was presented many times, across multiple days, but the *Shape-Shape* trial type was presented on 85% of the trials within each exposure to the procedure (i.e., each of the three remaining trial-types was presented on only 5% of the trials). Eventually, one might expect to observe the emergence of a STTDE for the

Shape-Shape trial-type because the coherence of this trial-type would increase relative to the others simply by dint of frequency of presentation.

Increasing the Complexity of the Model

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, a 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.[4] 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

[4] 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.

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-type, which reduces the tendency to pick "True" over "False", at least when compared to the low-fear participants. If the foregoing interpretation is correct it would explain why performance on this trial-type appears to predict actual approach towards a spider.

With that said, the question still remains why the *Spider-Approach* trial-type tends to produce an IRAP effect that is opposite in direction to that expected (choosing "True" more quickly than "False") across a normative sample of participants? Indeed, this weak/opposite effect has been observed in other studies using completely different stimuli (e.g., Kavanagh, Matthyssen, Barnes-Holmes, Barnes-Holmes, McEnteggart, & Vastano, 2019). Specifically, when performance on the two trial-types that require responding "False" during history-consistent blocks are compared with each other, the effect for the *negative-positive* trial-type is often weaker than for the *positive-negative* trial-type. How might we explain this difference, given that both trial types required the same RCI within blocks of trials? Once again, the DAARRE model may be of use here. If we examine Figure 2, it becomes apparent that the *Spider-Approach* trial type presents a *target* stimulus that coheres with the "True"

RCI in terms of its Cfunc properties (the *Pet-Avoid* trial-type presents a *target* that coheres with the "False" RCI). If we assume that the spatial contiguity between the target stimulus and the response option plays a role in determining the IRAP effect, the difference in trial-type effects observed here makes sense. More informally, participants may experience a "*Yes-No-No*" reaction to the *Pet-Avoid* trial type, but a "*No-Yes-No*" reaction to the *Spider-Approach* trial type, assuming that in general they read each IRAP trial from the top-down. If participants find it easier to select an RCI that is functionally similar to the target stimulus they have just observed, than an RCI that is functionally *dissimilar*, the weaker (or opposite) effect for the *Spider-Approach* trial type is readily predicted. We refer to this effect as the Dissonant Target Trial Type Effect (DTTTE; Finn & Barnes-Holmes, 2019; Kavanagh, et al).

Interpreting the Dynamics of Differential Trial Type Effects

In recognizing differential trial type effects, such as the STTDE and the DTTTE, we have begun to speculate that their presence (versus absence) could be important in revealing the relative dominance of Cfunc versus Crel properties of the stimuli presented within an IRAP. Critically, this insight could have important implications for developing a relatively precise experimental analysis of at least some of the behavioral processes that may be involved in more loosely defined middle-level concepts, such as fusion/defusion, which are common in the literature on acceptance and commitment therapy (see Barnes-Holmes, Hussey, McEnteggart, Barnes-Holmes, & Foody, 2016). Allow us to explain.

Consider an IRAP performance that reveals a large STTDE and DTTTE in an IRAP composed of positive and negative stimuli (see left-hand panel of Figure 3). Such a pattern of effects could be interpreted as indicating the relative dominance of the Cfunc properties of the stimuli over their Crel properties. That is, the orienting/evoking properties of the stimuli appear to have a relatively strong influence over the IRAP performance. In contrast, consider an IRAP performance in which the STTDE and DTTTE are relatively weak or absent and

each of the individual trial-type effects are quite similar (see right-hand panel of Figure 3). In this case, the Crel properties of the stimuli appear to be dominating the Cfunc properties. In other words, the coherence versus incoherence of the arbitrarily applicable relations between the label and target stimuli, within each trial-type, appear to be driving the IRAP performance. More informally, the presence of a STTDE and a DTTTE may indicate that a participant is strongly influenced by the orienting/evoking functions of the stimuli. Or in middle-level terms, the individual is *fused* with the psychological content of those stimuli. In contrast, if the STTDE and DTTTE are weak or absent, and the individual trial-type effects are quite similar, the participant is simply responding to the "cold" abstract relations among the stimuli. Or in middle-level terms, the individual is *defused* from the psychological content of the stimuli presented within the IRAP. Note that such a pattern indicates that the participant is still reacting to the semantic "meaning" of the stimuli, by showing response biases that reflect the coherence versus incoherence of the stimulus *relations*. Critically, however, the orienting/evoking functions of the stimuli appear to be relatively "distant"[5]

Insert Figure 3 Here

Conclusion

As noted above, the DAARRE model interpretations offered here are post-hoc and speculative, but critically they serve 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.

[5] If each of the individual trial type effects is close to zero, or are in a direction opposite to that expected based on the semantic meaning of the stimuli, it is difficult at the current time to interpret such a performance in terms of fusion versus defusion. That is, the stimuli do not appear to have the expected Crel properties (or "cold" semantic meanings) and thus the extent to which they may also possess the relevant Cfunc properties (or the "hot" attentional or emotional functions) is questionable. We should also be clear that, in our view, the relative dominance of Cfunc versus Crel properties in an IRAP does not provide a complete functional-analytic interpretation of the middle-level concepts of fusion/defusion. In general, middle-level concepts will remain at that level because they were not wrought *directly* from laboratory-based functional analyses. Thus we are simply suggesting that there may be some useful functional overlap between fusion/defusion and the relative dominance of Cfunc versus Crel control as revealed using the IRAP as a context for exploring the dynamics of AARRing.

For example, most participants may entail a "not" relation between spiders and approach but participants 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. And this effect may be further complicated by a general tendency to choose the RCI that coheres in terms of its Cfunc properties with the Cfunc properties of the target stimuli. Interestingly, in considering these complexities, and the relative strengths of specific effects, such as the STTDE and the DTTTE, we may develop a more experimentally grounded, bottom-up approach to tackling middle-level concepts such as fusion/defusion, hot versus cold cognitions, comprehensive distancing, and so forth. Indeed, recognizing these complexities, and their potentially wide-ranging implications, led us to integrate the MDML framework with the DAARRE model, yielding a hyper-dimensional multi-level framework (or HDML). Critically, this integration generated a new conceptual unit of analysis for RFT (i.e., the ROE), which stands for relating, orienting, and evoking. In the next section of the chapter, 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 analogs 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. Upon reflection, however, the property of entailment appears to be the key focus of

the MDML, 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 up-dated the MDML framework by integrating it with the DAARRE model, and we now refer to the MDML as the *H*DML, where the 'H' stands for 'hyper' (see Barnes-Holmes, 2018; Barnes-Holmes, et al., 2020). The term hyper is used because the integration does not simply involve adding additional dimensions to the MDML 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. The integration of the MDML framework and the DAARRE model is represented in Table 2.

Insert Table 2 Here

As can be seen in the table, 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.[6] In brief, *relating* refers to the myriad complex ways in which verbal humans can relate stimuli and events; *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 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

[6] 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 may 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.

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 will constantly 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. 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* particularly 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 human psychological events (see Gomes, Perez, de Almeida, Ribeiro, de Rose, & Barnes-Holmes, 2019, for a recent study that indicates that a

derived transformation of functions observed with an IRAP may be manipulated using a motivative contextual variable).

The Verbal Self and the ROE

In proposing the ROE as a new conceptual unit of analysis for RFT, it is important to emphasize that it is seen as inherent in the RFT concept of a verbal self. For RFT, it is axiomatic that without AARRing there would be no verbal self, and without a verbal self AARRing, at best, would be extremely constrained and limited. We assume, therefore, that *once a verbal self is established in the behavioral repertoire of an individual, it becomes an on-going 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 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 most trivial to the most fundamental of psychological events for (verbal) humans are seen as embedded in a constant and iterative daily cycle of ROEing.

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 psychological events are only made possible through the evolution of human language and our learning of a specific language through our on-going interactions with the verbal communities in which we reside from birth through to death.

The ROE and Verbal Self-reports

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 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. 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 nor do you 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 thus suggests that following the rule, rather than space-bar pressing *per se,* was reinforced.

In contrast, an *up-dated* 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. 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; Harte, Barnes-Holmes, Barnes-Holmes, & McEnteggart, 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. In other words, we 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, or something broadly similar, seem to be required even when verbal humans respond in very brief periods of time, as is the case with the IRAP.

Conclusions

The study of derived stimulus relations, as a vehicle for analyzing human language and cognition, commenced almost half a century ago with the seminal study by Sidman (1971) on stimulus equivalence relations as the basis for basic reading ability. In the intervening years much has been achieved in building out the basic concept of equivalence relations and in coming to appreciate the extent to which such relations do in fact provide a rich conceptual basis for developing a behavior-analytic account of human language and cognition. Relational frame theory, as articulated in the earliest writing in the area (Hayes, 1991; Hayes & Hayes, 1989) and in the seminal volume (Hayes, et al., 2001), and more recent works (e.g., Dymond & Roche, 2012; Hughes & Barnes-Holmes, 2016a, 2016b), provide one example of that on-going human-focused research program.

The current chapter provides a summary of a recent attempt to up-date RFT, as a modern behavior-analytic account of human language and cognition. In doing so, it can be seen that the theory now seems to require a greater focus on complex relational networking rather than on merely framing. Ironically, this focus was clearly present in its earliest exposition (Hayes & Hayes, 1989), in which it was used to provide a well-defined, functional-analytic treatment of rule-governed behavior as involving a network of relational frames. The

recent 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, appears to facilitate a renewed focus on relational networking. And as noted at the beginning of the current chapter, the original Kantorian flavor to RFT has reemerged for the current authors in conceptualizing complex relational networking as involving a field of behavioral (verbal) interactants. Allow us to explain.

In our view, the individual elements within any given relational network do not exist independently of each other; 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 definition of a psychological event and the psychological event is the field—they are one and the same "thing." There is no person (or verbal self) *contained* within the field; rather the person is treated as a constantly changing or actualizing field of verbal interactants.

Of course, the critical question emerges; does this of way of conceptualizing psychological events lead to improvements in behavioral prediction-and-influence, with precision, scope, and depth, relative to alternative ways of talking about psychology? 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 seems like an investment worth making. We are also aware that the increasing complexity and sophistication of the types of analyses that are being developed here may well appear daunting and extremely challenging. As we move forward, advanced mathematics, including techniques borrowed from other sciences, such as machine learning, combined with new and emerging technologies for gathering increasingly rich data sets, inside and outside of

the experimental laboratory, will likely be needed to explore more fully the behavioral dynamics of derived relational responding. But of course, such challenges are also exciting and provide evidence that the basic science of behavior analysis provides a highly fertile ground for future generations of researchers to grapple with the mysteries of human psychology inside a conceptual framework that remains functional-analytic-abstractive, naturalistic, and monistic. In any case, the material presented in the current chapter suggests that 50 years into the research story on derived stimulus relations, we are not at the end of the journey. Indeed, to paraphrase the mid-20th Century British Prime Minister, Winston Churchill, we are not even at the beginning of the end, but perhaps at the end of the beginning.

References

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

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

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., 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., 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., 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, Y., Hussey, I., McEnteggart, C., Barnes-Holmes, D., & Foody, M. (2016). Scientific ambition: The relationship between relational frame theory and middle-level terms in acceptance and commitment therapy. In R. D. Zettle, S. C. Hayes, D. Barnes-

Holmes, & A. Biglan (Eds.), The Wiley handbook of contextual behavioral science (pp. 365–882). West Sussex: Wiley.

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

Brownstein., A. J., & Shull, R. L. (1985). A rule for the use of the term, "Rule-Governed Behavior". *The Behavior Analyst*, 8, 265-267.

Dymond, S., & Roche, B. (2012). *Advances in relational frame theory: Research and application.* Oakland, CA: New Harbinger.

Finn M, & Barnes-Holmes D (2019). *Predicting-and-influencing patterns of arbitrarily applicable relational responding in individual performances in the Implicit Relational Assessment Procedure.* Paper presented at the Association for Contextual Behavioral World Conference, Dublin, Ireland.

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.

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

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.

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,* 743-763.

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

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.). *Dialogues on verbal behavior* (pp. 19-40). Reno, NV: Context Press.

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

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

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.

Kavanagh, D., Barnes-Holmes, Y., Barnes-Holmes, D., McEnteggart, C., Finn, M. (2018). Exploring differential trial-type effects and the impact of a read-aloud procedure on deictic relational responding on the IRAP. *The Psychological Record, 68,* 163-176.

Kavanagh, D., Matthyssen, N., Barnes-Holmes, Y., Barnes-Holmes, D., McEnteggart, C., & Vastano, R. (2019). Exploring picture of self and other in the IRAP: Reflecting on the emergence of differential trial-type effect. *International Journal of Psychology and Psychological Therapy, 19,* 323-336.

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.

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*, 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*, 387-398.

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.

Pinto, J. A. R., de Almeida, R. V., & Bortoloti, R. (2020). The stimulus' orienting function may play an important role in IRAP performance: Supportive evidence from an eye-tracking study of brands. *The Psychological Record*, https://doi.org/10.1007/s40732-020-00378-2.

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

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

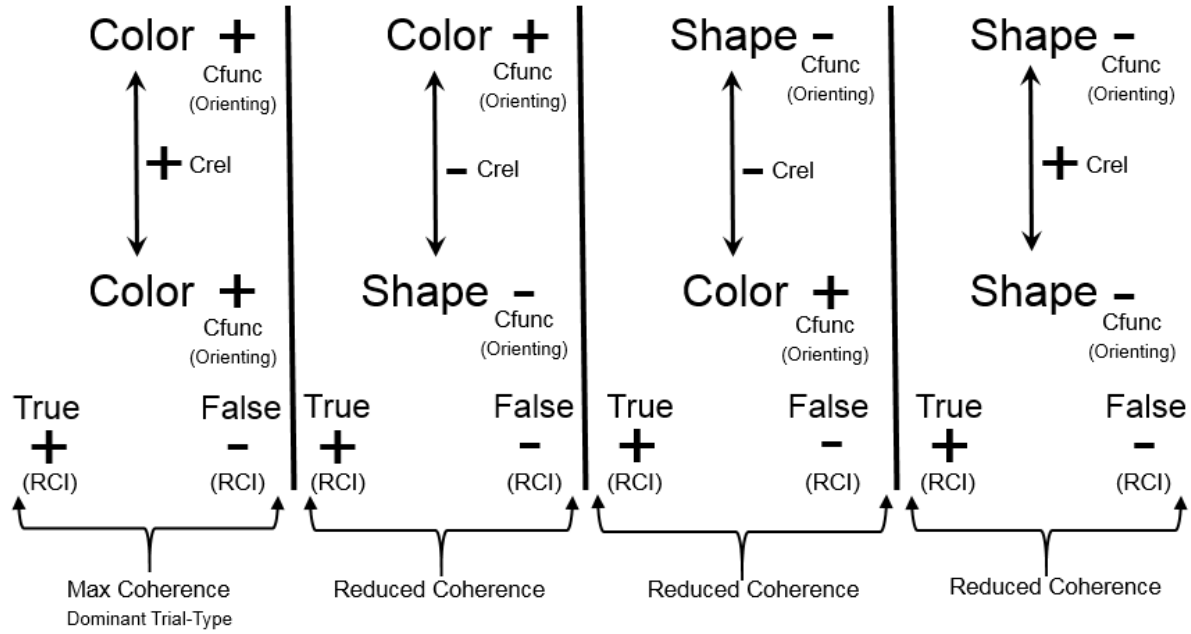

Figure 1. The DAARRE model as it applies to the "Shapes-and-Colors" IRAP. The positive and negative symbols refer to the relative positivity of the transformation-of-function property (Cfunc), for each label and target, the relative positivity of the entailment property (Crel) and the relative positivity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

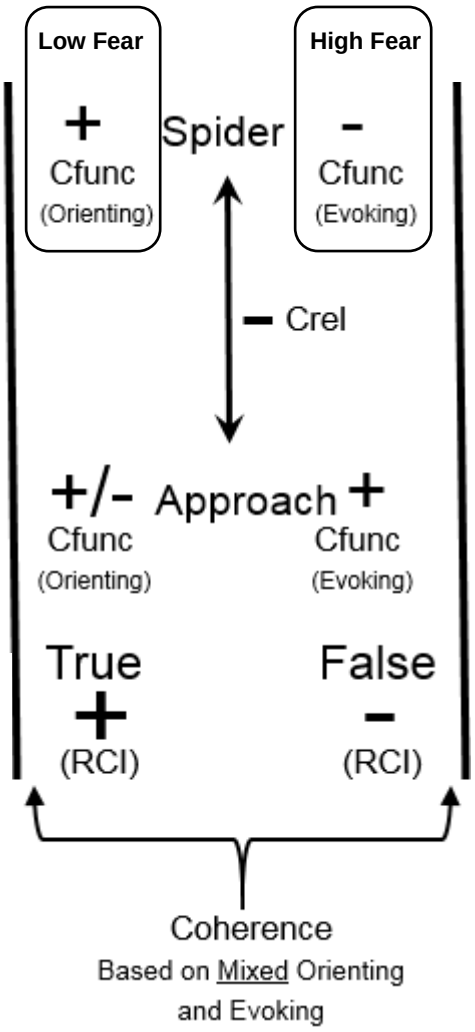

Figure 2. The DAARRE model as it applies to the Spider-Approach trial-type of the "Pets-and-Spiders" IRAP. The terms "Low Fear" and "High Fear" indicate the Cfuncs that are likely to dominate for individuals who are low (orienting) versus high (evoking) in spider fear. The assumption that the orienting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

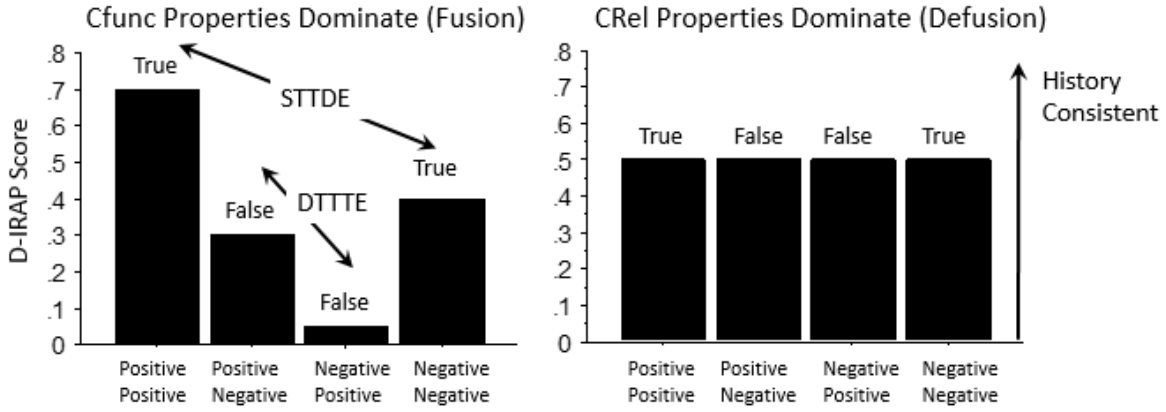

Figure 3. Interpreting the dynamics of differential trial type effects. Left panel: Hypothetical IRAP data illustrating the single trial type dominance effect (STTDE) and the dissonant target trial type effect (DTTTE). Right panel: Similar or "flat" trial type effects in which both the STTDE and the DTTTE are absent. The presence of the STTDE and/or the DTTTE may indicate "fusion" with the orienting/evoking functions of the stimuli presented within the IRAP; if these effects are relatively small or absent this may indicate "defusion" from the orienting/evoking functions (see text for details).

Table 2

*A **Hyper**-Dimensional Multi-Level (MDML) 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.*

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