Grappling with the Complexity of Behavioral Processes in Human Psychological Suffering: Some Potential Insights from Relational Frame Theory

Grappling with the Complexity of Behavioural Processes in Human Psychological Suffering: Some Potential Insights from Relational Frame Theory

Colin Harte[1,2,3], Dermot Barnes-Holmes[4], Julio C. de Rose[1,3], William F. Perez[2,3], & João H. de Almeida[3,5]

[1]Universidade Federal de São Carlos, São Carlos, Brazil

[2]Paradigma – Centro do Ciências e Tecnologia de Comportamento, São Paulo, Brazil

[3]Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino, São Carlos, Brazil

[4]Ulster University, Coleraine, Northern Ireland

[5]Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, Brazil

Author Note: Preparation of the current article was supported by grants from São Paulo Research Foundation (FAPESP; Grant 2019/24210-0 and Grant 2014/50909-8), and a grant from the Brazilian National Research Council (CNPq, Grant 465686/2014-1), to the National Institute of Science and Technology on Behaviour, Cognition, and Teaching (INCT-ECCE). The authors also thank the chairperson of INCT-ECCE, Deisy de Souza, for continuous support and encouragement. Correspondence for this article should be made to Colin Harte (colin.n.harte@gmail.com).

Abstract

Relational frame theory (RFT) has historically been considered the basic explanatory science behind acceptance and commitment therapy (ACT). However, some have argued that there has been an increasing separation between the two in recent years. The primary aim of the current article is to explore the extent to which RFT concepts, particularly those that have been proposed recently in the context of "up-dating" the theory, may be used to build stronger links between basic and applied behaviour analyses in which there is a shared language of relatively precise technical terms. As an example of this strategy, we outline RFT process-based experimental and conceptual analyses of the impact of one of the most widely used sets of interventions employed in the ACT literature, defusion. In addition, we suggest a potential experimental methodology for analysing the basic behavioural processes involved. Overall, the current article should be seen as part of a broader research agenda that aims to explore how RFT may be used to provide a functional-analytic abstractive treatment of the behavioural processes involved in human psychological suffering.

KEYWORDS: RELATIONAL FRAME THEORY; BEHAVIOURAL PROCESSES, PSYCHOLOGICAL SUFFERING; DEFUSION; ACCEPTANCE AND COMMITMENT THERAPY

Research on Acceptance and Commitment Therapy (ACT; Hayes et al., 1999, 2011), and mindfulness-based therapies generally, has grown exponentially in recent years (e.g., Gloster et al., 2020). Similarly, research on Relational Frame Theory (RFT; Hayes et al., 2001) has also grown considerably (e.g., Dymond et al., 2010; O'Connor et al., 2017), but certainly not at the same pace as research on ACT. One consequence of these differential growth patterns is that new concepts and theoretical middle-level terms have emerged in the ACT literature (e.g., defusion) that have limited value in conducting basic (functional) experimental analyses of human psychological suffering and distress (see Barnes-Holmes et al., 2016; see also Harte & Barnes-Holmes, 2022, for a discussion of the limitations of middle-level terms within experimental-analytic research).

In recent years, some authors have questioned the increasing separation between ACT and RFT (Barnes-Holmes et al., 2016). Part of the motivation for raising this question was the historical narrative that RFT provides the basic science underpinning ACT (Hayes et al., 1999). If this relationship is no longer in place, or has at least weakened to a considerable degree, then it seems important to meet that fact head-on and either seek to re-establish the relationship or abandon it. Indeed, addressing this issue seems particularly important given recent calls to focus on process-oriented single-subject research in the field (Hofmann & Hayes, 2019). In responding to a new-found emphasis on clinical processes at the level of the individual, RFT as a behaviour-analytic account of human language and cognition would appear to be well positioned to respond to this call. The current article outlines one example of a process-based experimental analysis of some aspects of human psychological suffering. Furthermore, the arguments we present here are part of a broader research agenda that is working towards reconnecting RFT with ACT, and clinical behaviour analysis more generally, in terms of basic processes, conceptualised and analysed empirically at the level of the individual.

Although the current article is part of a broader research agenda, it will focus on one particular concept that is widely used in the ACT literature, that of fusion/defusion. The former refers to being fused with psychological content, for instance, taking literally the narrative of one's own suffering; whereas the latter refers to any context in which the verbal functions sustained by the context of literality are undermined, thus creating a sense of distance from one's own suffering (i.e., undermining fusion; e.g., Blackledge, 2007). There is wide agreement in the literature that the concept is to be interpreted as a middle-level term. Such terms are seen as having more precision than folk psychological concepts but lack the *same level of* precision, scope and depth of technical terms (see Barnes-Holmes et al., 2016 and Hayes et al., 2012). As such, middle-level terms are designed primarily to orient researchers and clinicians towards a particular domain rather than providing a relatively precise functional-analytic definition of the relevant behavioural processes. With that said, it is assumed that middle-level terms aim to capture or refer to general or loose clusters of the key functional relations involved in the psychological events to which they refer. In the case of fusion/defusion, for example, it is argued that a cognitive defusion exercise or intervention does not seek to create new verbal stimulus relations, such as attempting to change the negative self-appraisal, "I am bad" to a positive appraisal, such as "I am good." Instead, a defusion intervention targets contextual variables involved (it is assumed) in providing the verbal stimulus relations with their fused, behaviour-controlling properties. Thus, for example, the detrimental psychological impact of the verbal self-appraisal as "Bad" may be undermined by changing the context in which the appraisal is typically produced. In the words of Assaz et al. (2018, p. 408),

> "rather than altering the content of verbalizations, cognitive defusion diminishes their impact on behaviour (Blackledge, 2007; Hayes et al., 2006; Hayes et al., 2012b). By decreasing the verbally conditioned

> functions of a stimulus, its directly conditioned functions as well as other stimuli…become more likely to exert influence over behaviour, possibly evoking more adaptive responses. This may reduce the experience of entanglement and fusion and enhance the experience of choice among different alternatives, increasing response flexibility (i.e., reducing rigidity)."

In reflecting upon the foregoing definition of the middle-level concept of fusion/defusion, it is certainly apparent that it seeks to specify the types of functional relations involved in the relevant behaviours. On balance, we would argue that there may be some value in using RFT to generate a relatively generic but precise specification of the core behavioural (verbal) processes involved in fusion/defusion. In making this argument, we are not suggesting that RFT will provide a point-to-point translation of the concept (fusion/defusion) itself (see Hayes et al., 2012); in effect, our aim here is not to render defusion a technical concept. Indeed, as we have argued previously, "such direct translation should not be expected because a primary purpose of basic or technical analyses is to generate verbal stimuli that control scientific behaviours (among basic and applied researchers, and practitioners) in new and unique ways; direct translation, by definition, cannot achieve this aim because one term is simply replaced with another and no substantive change in scientific or therapeutic behaviour is required or encouraged." (Harte & Barnes-Holmes, 2021a). Thus, our aim in the current article is simply to explore the extent to which RFT concepts, and particularly those that have been proposed recently in the context of "up-dating" the theory (e.g., Barnes-Holmes et al., 2017; Barnes-Holmes et al., 2020; Barnes-Holmes et al., 2021; Barnes-Holmes & Harte, 2022a; Harte & Barnes-Holmes, 2022), may be

used to provide a functional-analytic abstractive treatment[1] of the behavioural processes involved in human psychological suffering. The outcome should not be seen as an alternative account but one that aims to supplement what is already available in the ACT/RFT literature. Before proceeding with our suggested RFT treatment of fusion/defusion it seems useful to consider a recent attempt to deconstruct the concept (fusion/defusion) in terms of processes of change[2]. In doing so, we will draw on the work of Assaz et al. (2018), but fully recognize that this constitutes only one of many different interpretations that may be offered in the literature.

Four Potential Processes of Change Involved in Fusion/Defusion

There are many defusion techniques or exercises in the ACT literature, but Assaz et al. (2018) argued that four different basic processes of change are involved in ACT-based exercises that are used to target fusion/defusion. Specifically, the authors argued that it was possible to categorize different defusion interventions according to each intervention's key process of change. We will now provide a brief summary of these proposed processes of change along with an example of a defusion intervention that is suggested to operate through each process.

1. ***Respondent Extinction, Counterconditioning, or Inhibitory Learning***

Assaz et al. (2018) first argue that it may be useful to conceptualise ACT-based exposure interventions as working through respondent extinction. Specifically, interventions of this nature typically involve a stimulus being contacted repeatedly until the eliciting

[1] Functional analyses may simply involve identifying functional relations between or among two or more variables. Functional-analytic *abstractive* research, however, is generally guided by the *abstraction* of behavioural principles that serve to increase the prediction-and-influence of behaviour with precision, scope and depth.

[2] A process of change is a generic term that does not necessarily imply a behavioural process. For example, Hayes et al. (2020) have presented an extended evolutionary meta-model (EEMM) that refers to processes of change, which may or may not be defined in behaviour-analytic terms. As an example, the process of change referred to as *cognition* is exemplified by both "cognitive reappraisal" and "defusion" (Hofmann et al., 2021); the former is a traditional or mainstream CBT term, whereas the latter is more closely associated with ACT, RFT and the behaviour-analytic tradition.

functions of that stimulus are reduced. For example, the classic word repetition exercise involves asking the client to repeat a relatively innocuous word, such as ‘milk’, over and over until they experience the word ‘losing’ some of its meaning. The exercise is then typically repeated with a word that is therapeutically relevant. As such, the authors suggest that it may be possible to consider this as an instance of respondent extinction because repeatedly presenting the stimulus may decrease or undermine the strength of its functional properties (Tyron, 2005). Furthermore, the authors argue that other possibilities of change processes for exercises of this nature (exposure to diminish the eliciting functions of stimuli) may involve counterconditioning (one response is replaced by another) or inhibitory learning (new relations are created among stimuli to alter the functions of those stimuli; e.g., ‘fearful’ to ‘safe’).

2. Differential Reinforcement

Assaz et al. (2018) suggest that it may be useful to conceptualise ACT-based defusion exercises in which the link between thoughts and actions is ‘disrupted’ as involving differential reinforcement of alternative responses. Specifically, interventions of this nature typically involve a stimulus being contacted while emitting responses that are different from, or incompatible with, responses typically evoked by that stimulus. These responses are then reinforced by the clinician or by the client themselves through successfully engaging in client-valued actions. Consider, for example, the “thoughts and feelings aren’t causes” exercise in which the client is asked to verbally tell themselves that they cannot perform a simple activity (e.g., “I cannot walk around the room”) and then simultaneously perform that activity (i.e., walking around the room). In this way, therefore, the client emits a response that is incompatible with the response typically evoked by the stimulus (e.g., responding in accordance with the verbal statement), which is then reinforced initially by the clinician and perhaps later through successful contact with valued action.

3. Recontextualization of Thinking as a Narrative (Descriptive Autoclitic)

The authors (Assaz et al., 2018) argue that other defusion interventions may work through the use of descriptive autoclitics as a way of introducing alternative functional contexts in the presence of certain stimuli/events to diminish their transformation of function properties (e.g., the extent to which they evoke or control emotional reactions). That is, some exercises use descriptive autoclitics to build a narrative around the variables controlling behaviour in a given instance, typically describing thoughts as a narrative that may make sense to the individual but do not always correspond to direct experience. Consider, for example, the "having thoughts" exercise. During this exercise, clients are encouraged to notice that their thoughts are merely thoughts that they "have" (as opposed to prescriptions of reality; e.g., the thought "I am a bad person" is merely a thought and does not necessarily mean the individual is a bad person) and to notice their inherently arbitrary and often flawed nature.

4. Recontextualization of Thinking as Spatially Distant

Finally, Assaz and colleagues (2018) argue that one more process through which some defusion exercises operate uses spatial distancing to introduce alternative functional contexts in the presence of certain stimuli/events (thus diminishing their function transformation). The authors cite a rich literature on discounting (e.g., Rachlin, 2006) and self-distancing (e.g., Ayduk & Kross, 2008; Misckowski et al., 2012) in which recontextualising experiences from a distanced perspective is reported to often reduce stimulus control. As an example, consider the "leaves on a stream" exercise. During this exercise, clients are asked to imagine that they are sitting beside a stream watching leaves float by and to put each thought as they have it on a leaf on the stream, and to simply watch it float downstream. Thus, exercises in which thoughts or events are recontextualized for the client through spatial distancing are suggested to operate in this way. Indeed, as the authors

point out, the traditional centrality of this process to ACT is further evidenced by the original name for the therapy during its development, "comprehensive distancing."[3]

A Generic RFT-based Treatment of the Concept of Fusion/Defusion

In reflecting upon the foregoing treatment of the middle-level concept of fusion/defusion, it is clear that specifying the types of functional relations involved in the relevant behaviours is a key focus. At the same time, however, it seems to us that using recent advances in RFT to generate a relatively precise, generic specification of the *core* behavioural (verbal) process involved in fusion/defusion may also be of some value. As noted above, in advocating for such an approach we are not suggesting a point-to-point translation of the concept itself (fusion/defusion) will be provided by RFT; such a strategy seems to be more consistent with the four "processes-of-change" approach adopted by Assaz et al. (2018), which appealed largely to traditional behavioural processes. Of course, there is considerable value in the approach adopted by Assaz et al., and broadly similar interpretive exercises may well be of benefit to the research community. However, our aim here is simply to explore the extent to which RFT concepts, including relatively recent ones, may be useful in providing a functional-analytic abstractive account of the behavioural processes involved in human psychological suffering; in this specific case the behaviours involved in fusion/defusion.

RFT: Core Concepts

RFT is a modern functional contextual account of human language and cognition and is rooted in the Skinnerian tradition while drawing heavily on Murray Sidman's work on

[3] Following submission of the first version of the current article, Assaz and colleagues (2022) published another account of the potential processes of change involved in defusion, introducing other terms and concepts to the four presented in the 2018 article listed above. Some of these concepts appeal more directly to RFT but it could not be considered an attempt to interpret the behavioural processes involved in defusion solely in RFT terms (e.g., there are references to multiple pathways of change such as the 'selective attention pathway'). In contrast, the current article constitutes an attempt to develop a relatively precise account of the key verbal behavioural processes likely involved in fusion/defusion based purely in RFT terms, and derived in particular from recent conceptual and empirical developments within the theory.

stimulus equivalence (e.g., Sidman, 1971; Sidman & Tailby, 1982; see Sidman, 1994, for a book length treatment). According to RFT, the core unit of human language and cognition involves framing events relationally, or in other words, responding to one stimulus in terms of another. A relational frame is defined as a generalized pattern of relational responding possessing the properties of mutual entailment (if A = B, then B = A), combinatorial mutual entailment (if A = B, and B = C, then C = A and A = C), and the transformation of functions. The distinction between relational entailment and the transformation of functions is critically important in RFT because it distinguishes between the act of relating stimuli in an "abstract sense" from the impact of that relating on the functions of those stimuli. The classic demonstration of the transformation of functions thus involves establishing a frame of equivalence among three stimuli (A=B=C) and then establishing a specific function for one of the stimuli (e.g., pairing A with an unpleasant taste or smell) and then observing that the other stimuli within the frame also acquire that function in the absence of direct training (B and C acquire at least some of the unpleasant taste or smell functions).

A Critically Important Distinction Within the Theory

An important distinction is made in RFT between relational entailment and transformation of stimulus functions, specifying that each instance of relational framing occurs under two types of contextual control. One kind of contextual control (i.e., Crel) specifies the particular type of relation defining the relational response (e.g., coordination, opposition, etc.), while the other kind of contextual control (i.e., Cfunc) specifies the particular behavioural functions (e.g., approach, avoidance, etc.) that are transformed in accordance with the relational response. In any instance of relational framing as an act in context, the stimuli may possess entailing, or Crel properties, and transformation of function, or Cfunc, properties. Even a simple tendency to orient more strongly towards one stimulus rather than another in your visual field may be based on these types of contextual control. For

example, identifying the name of your favourite soccer team from a random list of team names may occur more quickly or strongly because it coordinates (i.e., it relates; a Crel property) with other stimuli that control strong orienting functions (e.g., the many highly familiar stimuli that constitute your favourite soccer team; a Cfunc property). The critically important distinction between Crel and Cfunc contextual control thus separates the two key properties of the stimuli that participate in relational frames (see Delabie et al., 2022; Finn & De Houwer, 2021; Perez et al., 2015, 2017, 2021, for relevant experimental support for this distinction).

The Implicit Relational Assessment Procedure

In attempting to measure and analyse the dynamics involved in relational framing and its associated properties, RFT research has increasingly employed the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006). The IRAP is a computer-based programme that requires participants to respond quickly and accurately to specific stimuli deemed to be either consistent or inconsistent with participants' pre-experimentally established learning histories. The general assumption is that, all things being equal, relational responding should be quicker and more accurate across blocks of trials that require relational responding that is consistent with participants learning histories (e.g., flowers are positive) than on blocks that require responding in a manner that is inconsistent with that history (e.g., flowers are negative). The primary datum from the IRAP is response latency, measured in milliseconds, and is defined as the time that elapses from the onset of stimulus presentation on each trial to the emission of a correct response.

Recent Advances in Analysing Response Patterns on the IRAP

In recent years, an unexpected pattern of IRAP results emerged when the stimuli presented within the procedure did not differ in an obvious way in terms of their valence or emotional properties. Most IRAP studies up until this recent research were assumed to differ

on this basis (e.g., in terms of race, body size, sexual preferences, etc., see Barnes-Holmes et al., 2020, for a recent summary of this work). Differences in the size of the trial-type effects could therefore often be explained in terms of the differential valences of the stimuli involved. However, when relatively non-valenced stimuli were inserted into the IRAP in recent research (see below), specific patterns emerged that could not be readily explained in terms of differential valence. One of these patterns has come to be termed a single-trial-type-dominance effect (STTDE; Finn et al., 2018). The pattern involves significant differences in magnitude between IRAP trial-type effects that share the same response options (e.g., see below) during history-consistent blocks of trials. For example, in the Finn et al. study, the authors employed an IRAP in which the label and target stimuli were comprised of shapes and colors. These label-target combinations yielded four individual trial-types; color-color, color-shape, shape-color, and shape-shape. In this case, the color-color and shape-shape trial-types required that participants press 'True' on history-consistent blocks and 'False' on history-inconsistent blocks; in contrast, the color-shape and shape-color trial-types required responding 'False' on history-consistent blocks and 'True' on history-inconsistent blocks. The standard assumption is that participants will respond more quickly on history consistent than on history inconsistent blocks.

Critically, consistently larger IRAP effects were observed for the trial that comprised colors as both labels and targets (i.e., the color-color trial-type) than the trial that comprised shapes as both labels and targets (i.e., the shape-shape trial-type). This difference was unexpected because these two trial-types required the same response option within each block of IRAP trials, and importantly did not differ in any obvious way in terms of their valence (i.e., no strong preference for colors over shapes). This, and similar STTDEs, could not be readily accounted for by existing models of IRAP performances (i.e., the REC model;

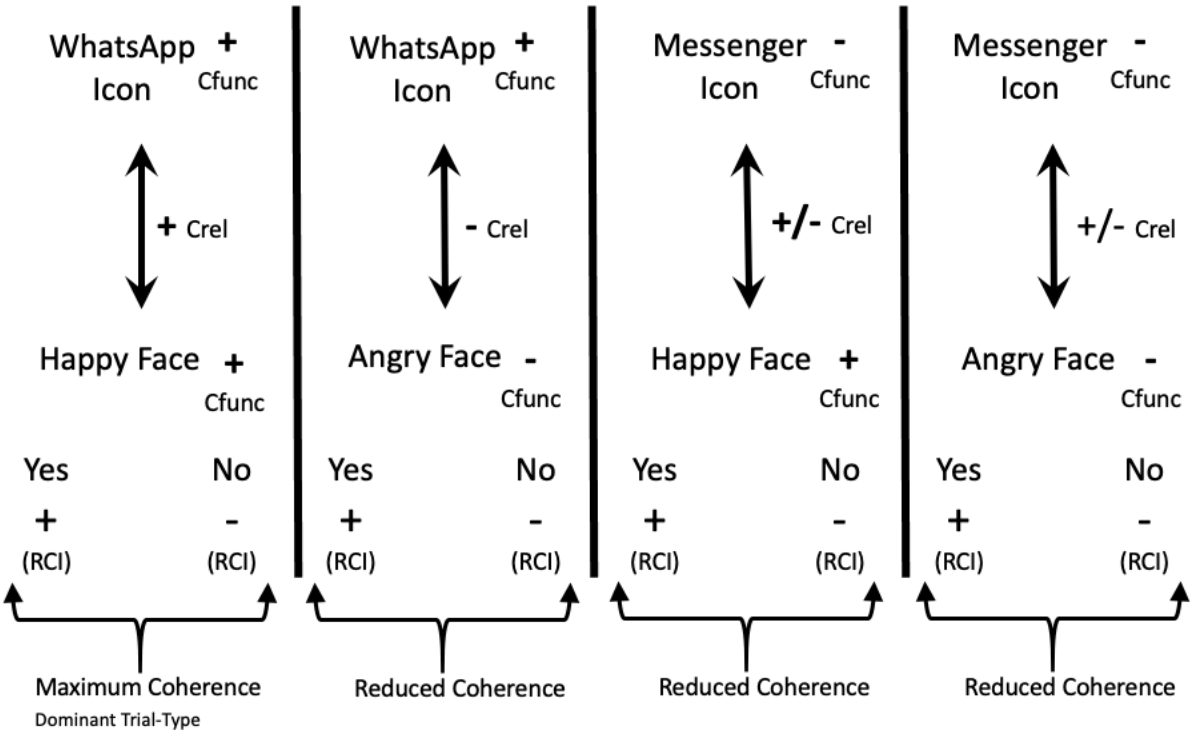
Barnes-Holmes et al., 2010), and so, a new model was proposed: the Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model (Finn et al., 2018).

The DAARRE model assumes that differential trial-type effects observed on the IRAP can be explained by the relative coherence between the Cfunc and Crel properties of the stimuli and response options employed across blocks of trials. In this context, response options such as "True" and "False" or "Yes" and "No" are termed relational coherence indicators (RCIs) given that they are often used to indicate coherence or incoherence between the label and target stimuli in a given IRAP (see Maloney & Barnes-Holmes, 2016, for a detailed treatment of RCIs). As an example, consider a recent study by Pinto et al. (2020) which reported a STTDE in an IRAP assessing preference for smartphone applications (WhatsApp versus Messenger). The WhatsApp and Messenger icons were employed as label stimuli, pictures of happy and angry faces as target stimuli, and the RCIs "Yes" and "No" as response options. Before completing the IRAP, participants were asked about their usage habits and preferences for apps. Following completion of the IRAP, participants took part in a visual search task using eye tracking technology to assess the degree to which the response biases produced on the IRAP corresponded with the stimuli towards which participants oriented more readily. An illustration of the DAARRE model as it applies to this particular IRAP is presented in Figure 1. Three key sources of behavioural control are identified by the model: (1) the relation between the label and target stimuli (Crels); (2) the orienting functions of the label and target stimuli (Cfuncs); and (3) the coherence functions of the RCIs (e.g., "Yes" and "No"). Interestingly, participant responding on the *WhatsApp Icon-Happy face* trial-type was significantly stronger (i.e., a larger D-IRAP score) than on all of the other three trial-types. That is, participants tended to respond "Yes" more quickly than "No" on the *WhatsApp-Happy Face* trial-type. On the *WhatsApp Icon-Angry face* trial-type, participants found it easier to respond "No" more quickly than "Yes", although this effect was

significantly weaker than the *WhatsApp Icon-Happy face* trial-type, and on the remaining messenger trial-types they responded with near equal latencies with "Yes" and "No."

Figure 1

The DAARRE model as it applies to the IRAP employed by Pinto et al. (2020)



Note. The '+' and '-' symbols are used to indicate the relative positivity of the Cfunc and Crel properties of the stimuli and the relational coherence indicators (RCI) relative to the other Cfuncs, Crels and RCIs in this IRAP.

In considering these results through the lens of the DAARRE model, it was assumed that the WhatsApp icon possessed relatively strong positive Cfunc properties relative to the Messenger icon. This assumption was made given that: all participants in the study Pinto et al. study (2020) reported using this application daily compared to less than half of participants reporting using the Messenger app daily; self-reported valence and arousal for WhatsApp were indicated as very positive or positive by most participants compared to

indifference for Messenger; and eye tracking data on the visual search task showed that participants consistently oriented toward the WhatsApp icon faster than the Messenger icon. The functions of these stimuli are thus labelled with a plus sign (+; for the WhatsApp icon) and minus sign (-; for the Messenger icon) in the figure to indicate their assumed Cfunc properties relative to one another. It was also assumed that the happy faces possessed positive Cfunc properties relative to the angry faces, also labelled with a + and - respectively in the figure (this was also assessed via the participants' self-reports). Furthermore, the Crel between the WhatsApp icon and happy faces is labelled with a + to indicate coherence, while the Crel between this icon and the angry faces is labelled with a '-' to indicate incoherence. The Crel on both Messenger trial-types is labelled with +/- to indicate a relative ambivalence in coherence, again all based on participant self-reports. Finally, the RCI "Yes" is labelled with a + to indicate coherence and the RCI "No" is labelled with a - to indicate incoherence (the assumption here is that affirmation is generally more positively valenced than negation; e.g., Nickerson, 1998).

The DAARRE model may thus explain the STTDE based on the extent to which the Cfunc, Crel, and RCI properties cohere across blocks of trials. The reader is again referred to Figure 1. First, note that each Cfunc and Crel property for stimuli in the WhatsApp Icon-Happy face trial-type are all labelled with + signs. During blocks of trials that require a "Yes" response (hereafter referred to as history-consistent blocks, inferred from participant self-reports and results of the visual search task), the model would consider this trial-type maximally coherent given the overlap between the functional properties of the stimuli and required RCI (all + signs). In contrast, during blocks of trials that require a "No" response (assumed to be history-inconsistent) on this trial-type, there is no coherence between the required RCI (labelled with a - sign) and the Cfunc/Crel properties of the stimuli. The

DAARRE model assumes that this contrast in coherence across history-consistent versus -inconsistent blocks of trials produces the strong IRAP effect observed on this trial-type.

Now consider the Messenger Icon-Angry face trial-type, which requires that participants choose the same RCI as the WhatsApp Icon-Happy face trial-type during history-consistent blocks. On this trial-type, the required RCI (plus sign) does not cohere with the Cfunc properties of the label and target stimuli (both minus signs), and its overlap with the Crel property is questionable given participants ambivalence here (the +/- sign). During history-inconsistent blocks of trials, the required RCI does cohere with the Cfunc properties of the stimuli (minus signs), but again its overlap with the Crel property is questionable. As such, the differences in coherence between history-consistent and history-inconsistent trials across these two trial-types are not equal, favoring the single-trial-type-dominance-effect. The pattern of effects produced on both other trial-types was also readily explained by the authors in the same way by the DAARRE model. What these analyses serve to illustrate, therefore, is that the DAARRE model may be used to provide relatively precise functional analyses of the variables influencing participant behaviour when completing an IRAP (see Barnes-Holmes & Harte, 2022b, for a summary of relevant work in this vein).

Using the Concept of Crel versus Cfunc Dominance to Develop Potential Analyses of Fusion/Defusion

Recently, a conceptual analysis of the ACT–based middle-level term, defusion, has been offered (Barnes-Holmes et al., 2021; Harte & Barnes-Holmes, 2021b) which *draws on* the relative dominance of Cfunc over Crel properties, as measured with the IRAP (the reader should note that we are *drawing on* these concepts, rather than using them to provide a point-to-point translation). To appreciate this analysis, consider again the STTDE as illustrated in Figure 2 (left-hand side). As argued previously, the DAARRE model interpretation of this effect (a relative dominance of trial-type 1 [e.g., color-color], over trial-type 4 [e.g., shape-

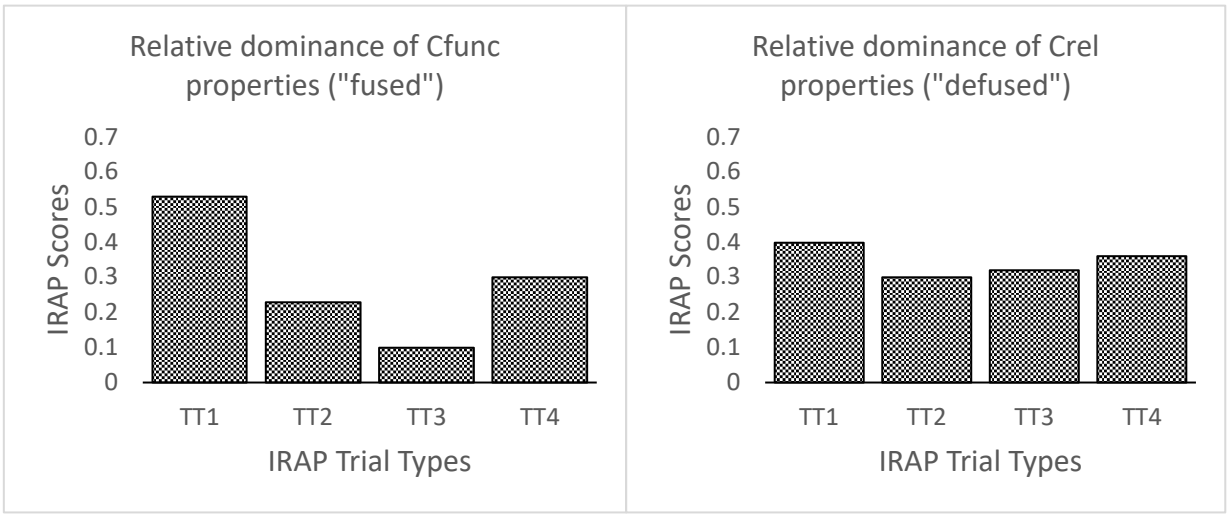
shape]) is taken to indicate that the Cfunc properties of the label and target stimuli strongly influence the IRAP performance. Or more informally, the Cfunc properties of those stimuli could be considered as showing relatively high fusion because they produce differential trial-type effects even though the same response option is required within blocks of trials. Now consider the right-hand side of Figure 2. In this case, all four IRAP trial-type effects are more or less even; critically, trial-type 1 does *not* dominate over trial-type 4. In this case, it appears that the Cfunc properties of the stimuli have a relatively limited impact on the IRAP performance. Less technically, the Cfunc properties are showing relatively low levels of "fusion" because the participant is simply relating the stimuli (responding to their Crel properties) without being unduly influenced by their non-relational (i.e., Cfunc) properties. In principle, therefore, these two patterns of responding on the IRAP might help to provide a relatively precise experimental analysis of the distinction between fusion and defusion.

While the foregoing conceptual analysis of fusion versus defusion provides a potential bridge between these middle-level ACT-based terms and basic experimental analyses within RFT, there is currently no *direct* empirical evidence available to support the proposed analysis. On balance, there is evidence that suggests that defusion exercises can impact upon IRAP effects (Kishita et al., 2014; Ritzert et al., 2015), although this research did not analyse patterns of IRAP effects at the level of the individual trial-type nor through the lens of the DAARRE model. One direction for future research, therefore, could involve exploring, testing, and developing this conceptual analysis. In broad terms, this would involve first replicating a STTDE effect (e.g., Figure 2 left-hand side) and then demonstrating that traditional defusion interventions employed in ACT reduce or "flatten" the STTDE. (e.g., Figure 2 right-hand side), preferably all at the single-participant level. As such, the core aim of the suggested research strategy would be to develop an increasingly precise process-based

account of defusion, and indeed clinical behaviour-analytic techniques more generally, in RFT terms.

Figure 2

Differential patterns of IRAP trial-type effects as potential models of fusion (Cfunc dominance) versus defusion (Crel dominance).



Note. The hypothetical data presented in the left panel illustrate the relative dominance of Cfunc control on the IRAP, suggested to be a potential model of "fused" responding (i.e., control of participant responding relatively dominated by the orienting and/or evoking properties of the stimuli). The hypothetical data presented in the right panel illustrate the relative dominance of Crel control on the IRAP, suggested to be a potential model of "defused" responding (i.e., control of participant responding relatively dominated by the purely relational properties between stimuli).

The research we are proposing here would involve implementing a series of defusion-like techniques in an attempt to undermine or reduce Cfunc dominance (i.e., fusion; Figure 2, left-hand panel) and increase Crel dominance (i.e., defusion, Figure 2, right-hand panel). Doing so would involve testing the argument that the basic behavioural (RFT) processes through which defusion works, at least in part, involve manipulating the Cfunc versus Crel properties of a specific relational network. We do recognize, of course, that the different interventions may not impact upon the relative dominance of Cfunc versus Crel control in

exactly the same manner and/or to the same extent. However, identifying and exploring such differences, experimentally, is the “sine qua non” of the research strategy that we are offering here because it will serve to provide opportunities to conduct increasingly precise experimental analyses of the behavioural processes involved when defusion techniques are employed in therapeutic contexts.

Factoring in Relational Complexity: A Hyper-Dimensional Multi-Level Framework

At this point it may seem a little simplistic to interpret fusion/defusion solely in terms of the relative dominance of Crel versus Cfunc control because, in our view, the domain is considerably more complex than this. Specifically, we would argue that the potential complexity involved in the behaviours associated with fusion/defusion may also be reflected in the context of a relatively new framework that has been proposed within RFT itself and may add to the analysis we have provided thus far. The framework is referred to as a hyper-dimensional, multi-level (HDML; Barnes-Holmes et al., 2020) space for conceptualising the complexities involved in arbitrarily applicable relational responding. A detailed exposition of this framework and a relevant graphical representation has been provided previously (e.g., Barnes-Holmes et al., 2020, 2021; Barnes-Holmes & Harte, 2022a), but a brief description of it will be outlined here. The framework divides relating into five increasingly complex levels; (i) mutual entailing, (ii) combinatorial entailing, (iii) relational networking, (iv) relating relations, and (v) relating relational networks. While the first two levels have already been described earlier in the article, we will briefly describe the final three levels.

Relational networking refers to combinations of combinatorially entailed relations (relational frames) into increasingly complex networks. This level of relating is typically appealed to when modelling and accounting for complex aspects of symbolic relating such as rule-following. For example, the rule “my dogs are very different in temperament, so feel free to pet one but not the other" is an instance of rule-following with only 2 relations within the

same network; coordination (between the words involved and events to which they refer) and difference (Zapparoli et al., 2021, p. 202). *Relating relations* involves relating one relational response to another relational response in some way. This level of relating is typically appealed to when modelling sophisticated symbolic relating such as that involved in analogy and metaphor. As a simple example, consider the simple analogy "red is to stop as green is to go." In this case, "red" and "stop" are coordinated through the Crel "is to", as are "green" and "go." Additionally, these two separate coordination relations are coordinated to one another via the Crel "as" (see Stewart & Barnes-Holmes, 2004, for a review). Finally, *relating relational networks* refers to, as the name suggests, the act of relating entire relational networks to other relational networks. This complex level of relating is likely involved when comparing and contrasting extended narratives and in complex problem solving. As Barnes-Holmes et al. (2017) argued, the following extended narrative that may occur in therapy likely involves multiple relational networks which are related to each other in order to support a clients claim that they are a failure: "I'm divorced, my kids don't talk to me, I'm still drinking too much, I think I'm just about to get fired from work, and when my mom died last year I was too drunk to attend the funeral" (p. 436).

The HDML framework also emphasises that relating at each level may vary dynamically along four dimensions: (i) coherence, (ii) complexity, (iii) derivation, and (iv) flexibility. In brief, *coherence* refers to the extent to which a pattern of relational responding is consistent with a previously established pattern. For example, after being told that "Monday comes before Tuesday", it is likely that the statement "Tuesday comes after Monday" would be considered coherent because it is consistent with what has been learned previously. *Complexity* refers to the level of detail involved in the relational responding, such as, for example, the types of relations involved. For example, coordination, which only involves one relation (if A = B then B = A), would be considered less complex than

comparison, which involves more than one relation (if A > B then B < A).[4] In addition, complexity may also involve increasing levels of contextual control (e.g., stimuli related on the basis of color alone, or stimuli related on the basis of both color *and* shape). *Derivation* refers to how frequently a pattern of relational responding has been emitted before. As a derived relational response is emitted, it is seen as decreasing in derivation, in that it develops its own history beyond the initial derivation. For example, if you are told that "Monday comes before Tuesday" and then derive for the first time that "Tuesday comes after Monday", this initial relational response would be considered high in derivation because it is derived directly from the original *Monday-before-Tuesday* relation. Every time the latter relation is derived thereafter, however, this response gradually acquires its own history, rendering it lower and lower in derivation. Finally, *flexibility* refers to how readily contextual variables can modify a pattern of derived relational responding. For example, if asked to respond with the wrong answer to "what day comes after Monday", the more readily an incorrect response is provided, the more flexible the relational response is deemed to be.

Each of the five levels intersects with each of the four dimensions to create a conceptual space for studying derived relational responding, highlighting multiple potential units of analysis that researchers can target. For example, imagine a researcher is interested in training and testing combinatorially entailed relations. By targeting this type of relation, the researcher has focused on a specific level of the HDML. They may then decide how many opportunities participants will have during testing to derive the targeted combinatorially entailed relation, thus invoking levels of *derivation*. In addition, the researcher may set a specific criterion for concluding that the relation is established (e.g., 12/15 trials correct), which could be interpreted as involving *coherence* (e.g., a larger number of correct trials may

[4] Note that the example above indicates that complexity can vary within a level of relational complexity, in this example a mutually entailed coordination relation is less complex than a mutually entailed comparison relation.

produce increased coherence). In addition, if the combinatorially entailed relation involved a comparison relation, for example, or indeed multiple stimulus relations in some way, *complexity* would be invoked, and if the malleability of this derived relation was assessed, for example, by reversing or changing the baseline training relations, then *flexibility* would also be targeted. In summary, therefore, the HDML framework provides a RFT-based conceptual space for highlighting and considering ways to target, manipulate, and analyse derived relational responding. [5]

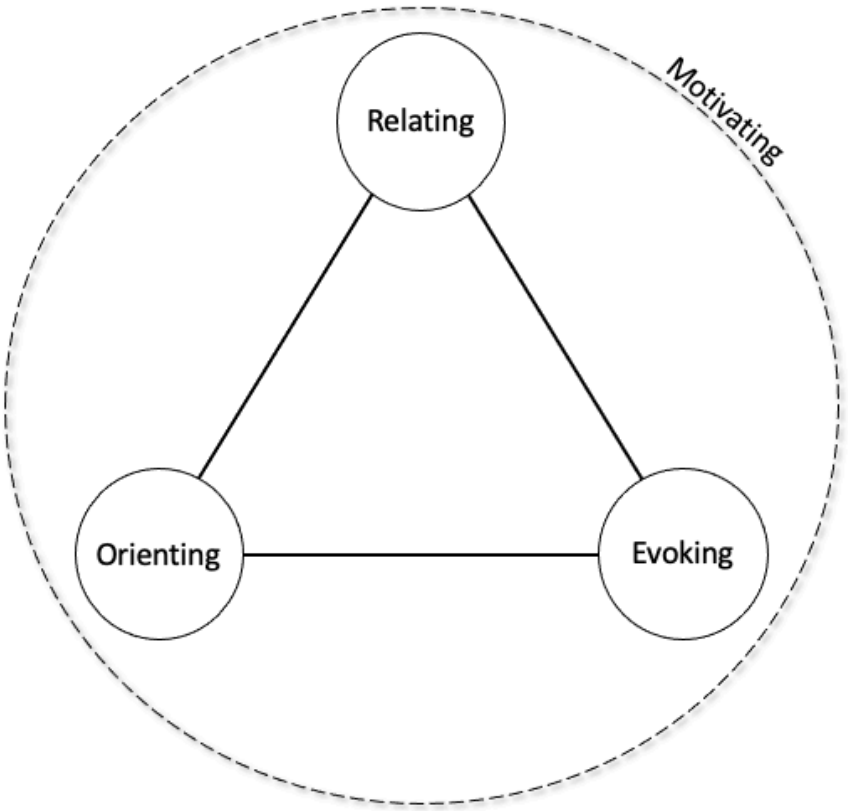
In addition, it is important to note that the framework is referred to as *hyper*- rather than *multi*-dimensional because it also highlights the central role of Cfunc properties in analysing derived relational responding. That is, in each instance of relating, at whatever level and along whichever dimension(s), the framework focuses on the orienting (i.e., the extent to which a stimulus is "noticed") and evoking (i.e., appetitive versus aversive) functions of stimuli that participate within stimulus relations and networks. Finally, the HDML also highlights the importance of the motivational context within which derived relational responding occurs in any given analysis. Considering again the hypothetical experimental example provided in the previous paragraph. The same researcher might also consider whether participants orient toward particular stimuli to a greater extent than others (e.g., using eye tracking technology; Pinto et al., 2020). They could also deem it important to assess or manipulate the extent to which the stimuli evoke appetitive versus aversive responses and how this controls, relative to orienting, participant responding (e.g., see Barnes-Holmes et al., 2020). And finally, differences in responding may also be explored as a function of motivational variables (e.g., Gomes et al., 2019).

[5] The units of analysis provided by the dimensions of the HDML have been generated through basic experimental analyses of behaviour. However, the analyses may also be clinically relevant. For example, someone presenting with death anxiety may engage in behaviour indicative of 'fusion', based on low derivation (e.g., many interactions in which death is coordinated with other aversive events) and low flexibility with thoughts around "death" (e.g., a history of interactions in which death is consistently related with other aversive events).

It has been proposed that a generic unit of analysis emerges from the HDML framework, which involves the dynamic interplay among relating, orienting and evoking, within a motivational context. This unit of analysis is referred to as the ROE-M (pronounced "roam"; Barnes-Holmes & Harte, 2022a; see Figure 3 for a graphic illustration). It is also important to highlight that contextual variables (e.g., antecedent and consequential stimuli) are inherent in this conceptualisation, via the four dimensions of the HDML from which the ROE-M emerges. For example, coherence highlights the extent to which a pattern of responding is consistent with a prior history of reinforcement; complexity incorporates types of contextual control involved in patterns of relational responding; derivation refers to how dependent current relational responding is upon an explicitly learned pattern of relational responding; and flexibility considers the extent to which a pattern of responding can be modified by current and historical contextual variables. Furthermore, highlighting the impact of motivating variables as a key aspect of the ROE-M analytic unit also explicitly emphasises the dynamic impact of contextual variables on the response (e.g., water as a reinforcer may differentially impact upon elements of the response based on how thirsty the individual is; e.g., Gomes et al. 2019). Overall, therefore, the consideration of these various contextual variables, which are built into the HDML framework, allow, in principle, for the prediction-and-influence of the ROE-M unit of analysis (see Barnes-Holmes and Harte, 2022, p. 251-252).

Figure 3

An illustration of the ROE-M unit of analysis



Note. Relating, orienting, and evoking are considered to be inseparable elements of the *response* unit as indicated by their connection in the triangle. While motivating is not a response per se, it is considered an ever changing context within which the elements of the response unit (the ROE) are influenced and co-determined. The ROE-M is thus presented as a dynamic, non-linear analytic response unit for conceptualising human psychological events.

The ROE-M and its Relevance to Conceptualising Defusion

Recent conceptual analyses have suggested that it may be useful to consider virtually all human psychological events as involving a dynamic and non-linear interaction among the elements of the ROE-M (Barnes-Holmes & Harte, 2022a; Barnes-Holmes et al., 2021). Critically, in the context of the current paper, it may be useful to consider if this conceptual behavioural unit may lend itself to the types of analyses we are proposing here. Although highly speculative, such an exercise may serve to illustrate the types of potential experimental analyses that could emerge through the lens of the HDML framework.

Before proceeding, we should emphasize that we are arguing that the ROE–M becomes the unit of analysis upon which contingencies impact. Given that the ROE-M, by definition, always involves arbitrarily applicable relational responding (the 'R' element of the unit), the ROE-M itself cannot be extracted from the wider HDML framework. Specifically, the framework provides the manipulable variables (i.e., dimensions) that in principle allow for the prediction and influence of the ROE-M analytic response unit itself. As such, the elements that comprise the ROE-M provide the behavioural processes through which change occurs. It thus follows that even well established procedures within behaviour analysis, such as extinction, operate on a verbal human in ways that differ from that of non-humans.[6] For example, the word-repetition "milk" exercise mentioned earlier may serve, through an extinction procedure, to undermine the control of orienting and evoking properties of the stimulus without impacting, in a substantive way, on the relating element of the ROE-M. More informally, following the milk exercise, the functions of the stimulus ("milk") may have dissipated to some degree (orienting and evoking properties) but the individual still knows what milk is and means (i.e., its relational properties); for example, the individual could point to a glass of milk when asked to do so. The intervention, therefore, may be seen as impacting differentially upon the elements of the ROE-M. As explained below, however, a detailed RFT interpretation of the defusion exercise involves more than simply targeting Cfunc over Crel properties. Specifically, for the exercise to have some therapeutic benefit it likely needs to be related to a similar exercise in which the stimulus is not "milk" but a word

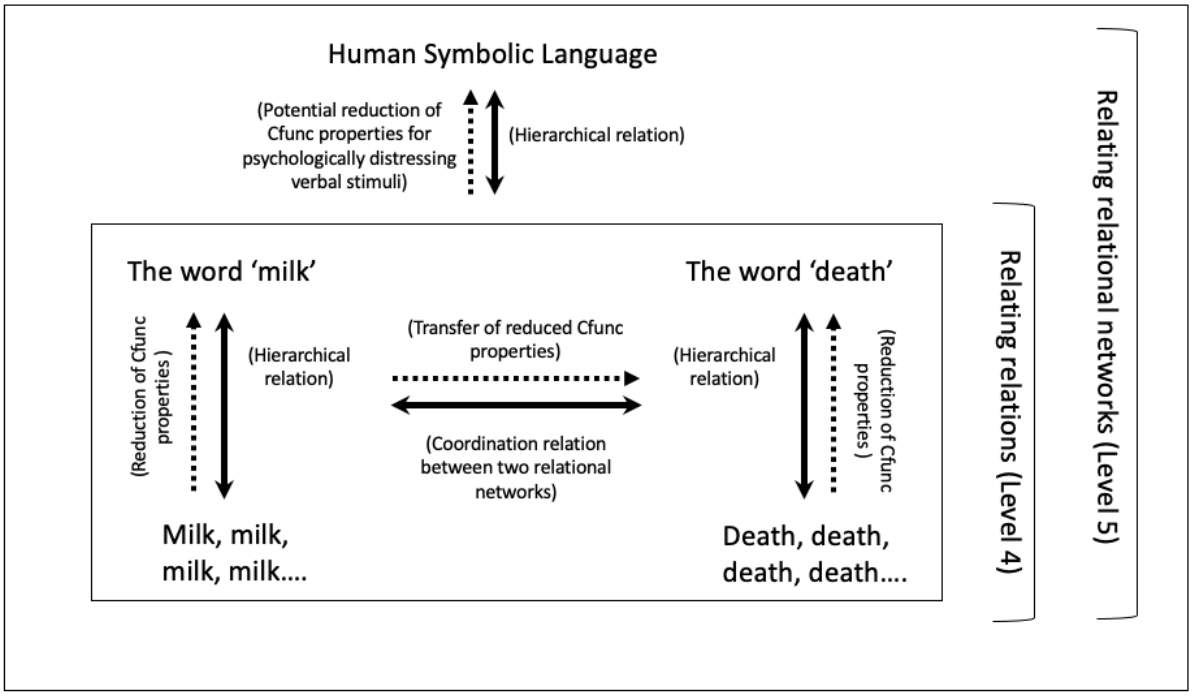
[6] From this perspective, relational framing alters the basic learning process; in other words, it changes how verbally competent human beings interact with the environment (e.g., Barnes-Holmes & Harte, 2022; Harte & Barnes-Holmes, 2021b). In adopting this view, it seems important to reconsider how behaviour-analytic processes are used to facilitate relatively fluid exchanges between basic and applied endeavours. In this regard, we would argue that beahvior analyses of human language and cognition should consider the ongoing stream of relating, orienting and evoking, and the impact of motivating variables, in attempting to predict and influence complex human behaviour.

that may produce a therapeutically relevant response in the client. For example, in the same way that the word “milk” may lose many of its Cfunc properties, so too might another word that is more relevant to the client’s suffering such as “death” (for an individual who is presenting with death anxiety). As such, the simple word-repetition exercise involves a great deal more than a reduction of Cfunc properties per se, but also involves relating the relational activity involved in the “milk” repetition exercise to the relational activity involved in a “death” repetition exercise (see Figure 4).

In this sense, the defusion exercise could be interpreted as involving relating relations in which the Cfunc-reducing properties for milk, through rapid repetition, are coordinated with “death” through rapid repetition to similarly reduce its Cfunc properties. But even this interpretation may be somewhat narrow or limited because for the defusion exercise to have its full therapeutic impact it needs to be related to language and thought itself. In effect, the therapist typically highlights that many words and thoughts may have a strong emotional impact on the client, but ultimately the emotional functions of these words and thoughts may literally “disappear” through a simple repetition exercise. Words and thoughts are not to be feared or avoided, therefore, but may be accepted and embraced for what they are, and not what they say they are (e.g., the word “death” is not death itself). Technically, therefore, the full defusion exercise involves the relating of relations to the superordinate relational network of human language itself (see Figure 4). As such, at the level of behavioural process, defusion seemingly involves a relatively complex cluster of relating relations and relating relational networks, rather than a simple reduction (or extinction) in Cfunc properties for a particular word or words.

Figure 4

Visual representation of the proposed relational activity involved during a hypothetical word repetition exercise



Note. The solid arrows indicate the type of relation (Crel) involved between stimuli and the broken lines indicate the changes in Cfunc properties that are suggested to occur. The repetition exercise is considered here to be in a hierarchical relation with the words involved (i.e., milk and death) because rapid repetition is one of many different things that can be done with a word.

Before proceeding, we should emphasise that the relating of relations in the way we have described above necessarily also involves deictic relational responding; that is, locating oneself verbally in space and time relative to others (I-You, Here-There, Now-Then; e.g., Barnes-Holmes, 2001; McHugh et al., 2004). Indeed, deictic relational responding has been seen as critical for the development of more complex forms of arbitrarily applicable relational responding itself, particularly relating relations (Barnes-Holmes & Harte, 2022a; see also Harte & Barnes-Holmes, 2021a, and Harte & Barnes-Holmes, 2021b). Providing further detail here is unnecessary in the context of the current article. Suffice to say that the relational

responding described in the examples provided in the current article should all be seen as involving deictic relational responding, particularly when relating relations is involved because this level of relational complexity seemingly requires relational responding to one's own relational responding.

Focusing on Orienting, Evoking and Motivation in the Analysis of Defusion

In outlining the foregoing process-based interpretation of a defusion exercise it is important to recognize that additional insights may be provided in terms of the ROE-M. For instance, in the above example we simply referred to a reduction in Cfunc properties without specifying the class of properties that might be targeted. During the course of therapy it may be useful for a therapist to explore the extent to which a client orients towards certain stimuli or events (i.e., is hypervigilant). For example, a death-anxious client might report that they are very focused on noticing or monitoring (i.e. orienting towards) their own heart rate, and when their heart rate begins to rise this evokes a strong anxiety response. The therapist might also explore the types of contexts in which such orienting and evoking responses occur. For example, the client might report that they become increasingly focused on their heart rate, and experience unexpected bursts in rate, whenever they have to attend certain social events, such as a work-related dinner. In this way, the orienting and evoking functions of the client's heart rate may be related to certain motivational variables (i.e., their death anxiety increases in social situations in which they may be judged negatively by work peers).

In terms of the interpretation of the defusion exercise considered previously, as the relating of relations and relating of relational networks, it may be useful to consider exactly what Cfunc properties are being targeted. For example, does the relating of relational networks undermine the orienting functions, the evoking functions, or both, and are motivational variables also being targeted? In this case, the relating relations process involved in the word-repetition exercises might help to undermine the orienting functions of

heart rate in that the verbal relations coordinated with heart-beat have lost some of their meaning. Or more informally, if the client has the thought “my heart is really beating fast” the relating relations process may undermine its problematic Cfunc property, such that the client does not then continue to monitor the rate for potential signs of a heart attack. Or in other words, the client thinks “the thought that a racing heart signals a heart attack is just another thought”. In addition, the relating of relations may also undermine the evoking functions of a rapid heartbeat if the client fails to ignore their racing heart. In effect, if the client actually begins to monitor their beating heart and starts to become anxious with thoughts such as “I am going to have a heart attack” the relating of relations outlined above may help to “remind the client” that these are just words or thoughts, thus reducing their Cfunc properties.

It is important to emphasise that we have used the term ‘targeting’ when referring to relating (AARR), orienting, evoking, or motivating. In effect, we are arguing that one or more of these elements may be targeted individually in any given analysis or intervention (experimental or clinical). However, it is worth reiterating that these elements are inseparable within the ROE-M as a non-linear, dynamic behavioural unit. Thus, while specifically targeting one over another may be deemed relatively more or less important in a given analysis, changes in one of these elements may produce changes in the other elements. In the example provided previously, therefore, a defusion intervention employed to undermine orienting and/or evoking functions may well produce changes in the relating and motivating elements of the response unit.

In any case, at the level of process, it could be argued that most if not all defusion exercises work in broadly the same way. For example, the floating leaves exercise, which involves the client placing each thought they have on a leaf as it floats down a stream, could be seen as reducing the Cfunc, but not the Crel, properties of individual thoughts.

Technically, the Cfunc properties of the thoughts may be reduced when they are placed in the context of a floating leaf, which disappears downstream. Indeed, the floating leaf may have multiple Cfunc and Crel functions in that the leaf fades away and its flow cannot be controlled, etc. Thus, a therapist may attempt to evoke particular Cfunc properties based on Crel control. For example, the therapist might emphasise the temporary and flowing properties of the leaf during the exercise. This may then evoke the Cfunc properties of 'feeling' or 'sensing' the disappearing and flowing properties of the leaf as it flows downstream. If the client is then invited to engage in this type of exercise with potentially problematic thoughts, such as "I am going to have a heart attack" then this could be interpreted as an instance of relating relations (i.e., benign thought on a leaf is related to problematic thought on a leaf). As suggested previously, the defusion exercise would also likely involve the relating of relations to the superordinate relational network of human language itself, such that the Cfunc properties of any thought may be reduced in contexts in which they are problematic for the client in their natural environment. And similar to the previous word repetition example, the therapist and client may focus on the orienting and evoking properties of particular thoughts and the motivational contexts in which they are likely to arise.[7]

Linking Technical Interpretations to Experimental Analyses

In providing RFT-based technical interpretations of defusion, it seems wise to ask what might be gained in doing so. In our view, the main advantage is that the concepts we have employed here are relatively precise, such as the levels and dimensions of the HDML, the transformation of functions, and the Crel and Cfunc properties of stimuli (by precise, we mean that such terms have been employed usefully in the experimental analyses of behaviour

[7] The reader should note that conceptualising clinical cases and formulating intervention strategies through the lens of the ROE (the response unit without the motivational context) typically characterises what has been labelled, Process-Based Behaviour Therapy (PBBT™ ; https://perspectivesireland.ie/pbbt/). Note, however, that the clinical interpretations outlined above are not formal PBBT™ derived analyses.

in the basic research laboratory). A relevant example is provided earlier in the current article in which the relative dominance of Cfunc versus Crel properties may be manipulated using the IRAP as a measure. Perhaps this basic idea could be further explored by examining the impact of other variables. For example, does Cfunc dominance increase with lower levels of derivation or higher coherence and/or complexity? How do these changes impact upon the flexibility of the network targeted? Could experiments be designed in which orienting or evoking functions are targeted individually to assess their impact on reducing Cfunc dominance? Perhaps various levels of the HDML could also be targeted. For example, recent studies have begun to explore the impact of targeting critical versus non-critical parts of a relational network in terms of its control over rule-governed behaviour (e.g., Bern et al., 2021; Harte, Barnes-Holmes, Barnes-Holmes, et al., 2021; Harte, Barnes-Holmes, Moreira, et al., 2021).[8] As another example, it may be important to explore the impact of one or more of the dimensions of the HDML on deictic relational responding (i.e., I/YOU, HERE/THERE, and NOW/THEN relations) and the extent to which such responding relies on the relating of relations and relating relational networks (e.g., see Harte & Barnes-Holmes, 2021a, for a relevant preliminary technical analysis). The purpose of such experimental analyses should not be seen as an end in itself, but as part of a research agenda that aims to build stronger links between basic and applied behaviour analyses, in which there is a shared language of relatively precise technical terms.

Conclusion

[8] In referring to the concept of rule-governed behaviour, we would argue that this concept itself lacks the level of technical precision we are seeking to develop here, both conceptually and experimentally (see Harte & Barnes-Holmes, 2022, for a detailed exposition of this argument). For this reason, we have refrained from drawing on concepts, such as broad classes of rule-governed behaviour (including pliance, tracking and augmenting) and broad classes of the concept of self (e.g., self-as-process and self-as-context). In doing so, we are not suggesting that such concepts have not been useful and may continue to be of some value. However, our strategy here is to move beyond an over-reliance on such middle-level concepts in continuing to develop an ever more sophisticated science of behaviour that can genuinely step up to the challenge of the human condition.

Overall, the arguments we have presented here aim to begin to build clearer links between RFT with clinical application, by providing RFT process-based conceptual analyses, which are clearly rooted in recent experimental research. As an example of this strategy, we have suggested a potential methodology for analysing the basic behavioural processes involved in one of the most widely used sets of interventions employed in ACT. Ultimately, however, the utility of these conceptual analyses and the concepts we have suggested will not be decided by one article or research group, but by the extent to which they produce a rich programme of experimental research and application. Nonetheless, we believe that developing a range of technical interpretations, such as those that we have attempted to provide here, and those provided by others (e.g., Assaz et al. 2018, 2022), will be beneficial to the field. The science itself will then select and retain the analyses and concepts that serve to most accurately control the behaviour of the scientist and practitioner in the development of a more refined account of human language and cognition, and its role in human psychological suffering.

Statements and Declarations

The authors declare that they have no conflicts of interest.

References

Assaz, D.A., Roche, B., Kanter, J.W., & Oshiro, C.K.B. (2018). Cognitive defusion in Acceptance and Commitment Therapy: What are the basic processes of change. *The Psychological Record, 68*, 405-418. https://doi.org/10.1007/s40732-017-0254-z

Assaz, D.A., Tyndall, I., Oshiro, C.K.B., & Roche, B. (2022). ~ process-based analysis of cognitive defusion in acceptance and commitment therapy. *Behaviour Therapy.* Advanced Online Publication. https://doi.org/10.1016/j.beth.2022.06.003

Ayduk, Ö., & Kross, E. (2008). Enhancing the pace of recovery: Self-distanced analysis of negative experiences reduces blood pressure reactivity. *Psychological Science, 19,* 229–231. https://doi.org/10.1111/j.1467-9280.2008.02073.x

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

Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2020). Updating RFT (more field than frame) and its implications for process-based therapy. *The Psychological Record, 70,* 605-624. https://doi.org/10.1007/s40732-019-00372-3

Barnes-Holmes, D., Barnes-Holmes, Y., McEnteggart, C., & Harte, C. (2021). Back to the future with an up-dated version of RFT: More field than frame? *Perspectivas em Análise do Comportamento, 12*(1). https://doi.org/10.18761/PAC.2021.v12RFT.03

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, & Stewart, I. (2006). Do you really know what you believe? Developing the IRAP as a direct measure of implicit beliefs. *The Irish Psychologist,* (32)7, 169-177.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the implicit relational assessment procedure (IRAP) and relational elaboration and

coherence (REC) model. *The Psychological Record, 60,* 527-542. https://doi.org/10.1007/BF03395726

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

Barnes-Holmes, D. & Harte, C. (2022a). Relational frame theory 20 years on: The Odysseus voyage and beyond. *Journal of the Experimental Analysis of Behaviour, 117*(2), 240-266. https://doi.org/10.1002/jeab.733

Barnes-Holmes, D. & Harte, C. (2022b). The IRAP as a measure of implicit cognition: A case of Frankenstein's monster. *Perspectives on Behaviour Science, 45,* 559-578. https://doi.org/10.1007/s40614-022-00352-z

Barnes-Holmes, D., Harte, C., & McEnteggart, C. (2020). Implicit cognition and social behaviour. In M. Fryling, R.A. Rehfeldt, J. Tarbox, & L.J. Hayes (Eds), *Applied Behaviour Analysis of Language and Cognition* (pp. 264-280). Context Press.

Barnes-Holmes, Y., Hussey, I., McEnteggart, C., Barnes-Holmes, Y., & Foody, M. (2016). Scientific ambition: The relationship between relational frame theory and middle-level terms in acceptance commitment therapy. In R.D. Zettle, S.C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds), *The Wiley Handbook of Contextual Behavioural Science* (pp 365-382), West Sussex, UK: Wiley-Blackwell.

Bern, R., Persdotter, T., Harte, C., & Barnes-Holmes, D. (2021). Relational coherence and persistent rule-following: The impact of targeting coherence in a 'non-critical' component of a relational network. *The Psychological Record, 71,* 279-290. https://doi.org/10.1007/s40732-020-00414-1

Blackledge, J.T. (2007). Disrupting verbal processes: Cognitive defusion in acceptance and commitment therapy and other mindfulness-based therapies. *The Psychological Record, 57*(4), 555-576. https://doi.org/10.1007/BF03395595

Delabie, M., Cummins, J., Finn, M., & De Houwer, J. (2022). Differential Crel and Cfunc acquisition through stimulus pairing. *Journal of Contextual Behavioural Science, 24*, 112-119. https://doi.org/10.1016/j.jcbs.2022.03.012

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

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*(1), 11–25. https://doi.org/10.1007/s40732-017-0262-z

Finn, M. & De Houwer, J. (2021). The selective action of Cfunc control. *Journal of the Experimental Analysis of Behaviour, 116*(3), 314-331. https://doi.org/10.1002/jeab.717

Gloster, A. T., Walder, N., Levin, M. E., Twohig, M. P., & Karekla, M. (2020). The empirical status of acceptance and commitment therapy: A review of meta-analyses. *Journal of Contextual Behavioural Science, 18*, 181-192. https://doi.org/10.1016/j.jcbs.2020.09.009

Gomes, C., Perez, W., de Almeida, J., Ribeiro, A., de Rose, J., & Barnes-Holmes, D. (2020). Assessing a derived transformation of functions using the implicit relational assessment procedure under three motivative conditions. *The Psychological Record, 69,* 487–497. https://doi.org/10.1007/s40732-019-00353-6

Harte, C. & Barnes-Holmes, D. (2021a, March 29). Wherever I "ROE-M", there I am: An RFT (technical) analysis of the verbal self and altered states of consciousness.

https://science.abainternational.org/wherever-i-roe-m-there-i-am-an-rft-technical-account-of-the-verbal-self-and-altered-states-of-consciousness/louise-mchughucd-ie/

Harte, C. & Barnes-Holmes, D. (2021b). A primer on relational frame theory (RFT). In M.P. Twohig, M.E. Levin, & J.M. Peterson (Eds.). *The Oxford Handbook of Acceptance and Commitment Therapy.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197550076.013.4

Harte, C. & Barnes-Holmes, D. (2022). The status of rule-governed behaviour as pliance, tracking and augmenting within relational frame theory: Middle-level rather than technical terms. *The Psychological Record, 72,* 145-158. https://doi.org/10.1007/s40732-021-00458-x

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2021). Exploring the impact of coherence (through the presence versus absence of feedback) and levels of derivation on persistent rule-following. *Learning and Behaviour*, *49,* 222-239. https://doi.org/10.3758/s13420-020-00438-1

Harte, C., Barnes-Holmes, D., Moreira, M., de Almeida, J.H., Aparecida-Passarelli, D., & de Rose, J.C. (2021). Exploring a Training IRAP as a single participant context for analysing reversed derived relations and persistent rule-following. *Journal of the Experimental Analysis of Behaviour, 115*(2), 460-480. https://doi.org/10.1002/jeab.671

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

Hayes, S.C., Barnes-Holmes, D., & Wilson, K. (2012). Contextual behavioural science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioural Science, 1*(1-2), 1-16. https://doi.org/10.1016/j.jcbs.2012.09.004

Hayes, S.C., Hofmann, S.G., & Ciarrochi, J. (2020). A process-based approach to psychological diagnosis and treatment: The conceptual and treatment utility of an extended evolutionary meta model. *Clinical Psychology Review, 82,* 101908. https://doi.org/10.1016/j.cpr.2020.101908

Hayes, S. C., Strosahl, K., & Wilson, K.G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behaviour change.* New York: Guilford Press.

Hayes, S. C., Strosahl, K., & Wilson, K.G. (2011). *Acceptance and Commitment Therapy: The Process and Practice of Mindful Change* (2nd ed.). New York: Guilford Press.

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science, 7*(1), 37-50. https://doi.org/10.1177/2167702618772296

Hofmann, S.G., Hayes, S.C., & Lorscheid, D.N. (2021). *Learning Process-Based Therapy: A Skills Training Manual for Targeting the Core Processes of Psychological Change in Clinical Practice.* California: Context Press.

Kishita, N., Muto, T., Ohtsuki, T., & Barnes-Holmes, D. (2014). Measuring the effect of cognitive defusion using the Implicit Relational Assessment Procedure: An experimental analysis with a highly socially anxious sample. *Journal of Contextual Behavioural Science, 3*(1), 8-15. https://doi.org/10.1016/j.jcbs.2013.12.001

Maloney, E., & Barnes-Holmes, D. (2016). Exploring the behavioural dynamics of the implicit relational assessment procedure: The role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record, 66,* 395–403. http://doi.org/10.1007/s40732-016-0180-5

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *The Psychological Record, 54*, 115-144. http://doi.org/10.1007/BF03395465

Mischkowski, D., Kross, E., & Bushman, B. J. (2012). Flies on the wall are less aggressive: Self-distancing "in the heat of the moment" reduces aggressive thoughts, angry feelings and aggressive behaviour. *Journal of Experimental Social Psychology, 48*, 1187–1191. https://doi.org/10.1016/j.jesp.2012.03.012

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. http://doi.org/10.1037/1089-2680.2.2.175

O'Connor, M., Farrell, L., Munnelly, A., & McHugh, L. (2017). Citation analysis of relational frame theory: 2009–2016. *Journal of Contextual Behavioural Science, 6*(2), 152–158. https://doi.org/10.1016/j.jcbs.2017.04.009

Perez, W. F., Fidalgo, A. P., Kovac, R., & Nico, Y. C. (2015). The transfer of cfunc contextual control through equivalence relations. *Journal of the Experimental Analysis of Behaviour, 103*(3), 511–523. https://doi. org/10.1002/jeab.150

Perez, W. F., Kovac, R., Nico, Y. C., Caro, D. M., Fidalgo, A. P., Linares, I., de Almeida, J. H., & de Rose, J. C. (2017). The transfer of Crel contextual control (same, opposite, less than, more than) through equivalence relations. *Journal of the Experimental Analysis of Behaviour, 108*(3), 318-334. https://doi. org/10.1002/jeab.284

Perez, W. F., de Azevedo, S. P., Gomes, C. T., & Vichi, C. (2021). Equivalence relations and the contextual control of multiple derived stimulus functions. *Journal of the Experimental Analysis of Behaviour, 115*(1), 405–420. https://doi.org/10.1002/jeab.649

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, 70,* 257–266. http://doi.org/10.1007/s40732-020-00378-2

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behaviour, 85*, 425–435. http://doi.org/10.1901/jeab.2006.85-05

Ritzert, T.R., Forsyth, J.P., Berghoff, C.R., Barnes-Holmes, D., & Nicholson, E. (2015). The impact of a cognitive defusion intervention on behavioural and psychological flexibility: An experimental evaluation in a spider fearful non-clinical sample. *Journal of Contextual Behavioural Science, 4*(2), 112-120. http://doi.org/10.1016/j.jcbs.2015.04.001

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

Sidman, M. (1994). *Equivalence relations and behaviour: 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 Behaviour, 37*(1), 5-22. http://doi.org/10.1901/jeab.1982.37-5

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

Tryon, W. W. (2005). Possible mechanisms for why desensitization and exposure therapy work. *Clinical Psychology Review, 25*(1)*,* 67–95. http://doi.org/10.1016/j.cpr.2004.08.005

Zapparoli, H.B., Marin, R., & Harte, C. (2021). Rule-governed behaviour: An ongoing RFT-based operant analysis. *Perspectivas em Análise do Comportamento, 12*(1), 197-213. https://doi.org/10.18761/PAC.2021.v12.RFT.09