# A process-based analysis of cognitive defusion in Acceptance and Commitment Therapy

Daniel A. Assaz[1]

Ian Tyndall[2]

Claudia K. B. Oshiro[1]

Bryan Roche[3]

1. Department of Clinical Psychology, University of São Paulo, Brazil
2. Department of Psychology, University of Chichester, United Kingdom
3. Department of Psychology, Maynooth University

Author note

Correspondence concerning this article should be addressed to Daniel Afonso Assaz. Institute of Psychology, University of São Paulo, São Paulo, Brazil, 05508-030. Email: daniel.assaz@alumni.usp.br.

**Abstract**

Cognitive defusion is among the main components of Acceptance and Commitment Therapy (ACT), a contextual behavioral approach to psychotherapy. Defusion serves as a middle-level term, and, as such, may be useful for applying and disseminating behavior science, despite its lower precision. However, some authors argue that for middle-level terms in psychotherapy to be useful to clinicians, they need to be clearly linked to basic behavioral concepts, with higher precision; and that this is not currently the case with defusion. Our objective is to increase the pragmatic utility of the concept of "cognitive defusion" by providing a more nuanced, multi-faceted and process-based definition of the term. In order to do this, we surveyed the ACT literature regarding defusion and critically examined it through the lens of conceptual analysis. This culminated in a revised and updated conceptualization of defusion in terms of its relationship to basic behavioral concepts, in which defusion is an outcome that may be achieved through different processes.

**Key-words:** Acceptance and commitment therapy, ACT, cognitive defusion, middle-level terms, behavioral processes

Acceptance and Commitment Therapy (ACT) is a contextual behavioral approach to psychotherapy that has demonstrated its efficacy across a wide range of mental health conditions, including anxiety, depression, substance abuse and chronic pain (Gloster et al., 2020). Since the early 2000's, ACT's therapeutic approach has been commonly presented through the Psychological Flexibility Model[1]. It is comprised of six components: acceptance, cognitive defusion, self-as-context, contact with the present moment, values and committed action (Hayes & Strosahl, 2004).

All of these are regarded as middle-level terms, concepts that intentionally forsake high precision[2] in favor of pragmatic utility in applying and disseminating behavioral science (Hayes et al., 2012; Vilardaga et al., 2009). Currently, middle-level terms are recurrently used and encouraged within Contextual Behavioral Science (CBS), a strategical and scientific approach to behavior science that encompasses ACT (Hayes et al., 2012).

However, relying on middle-level terms uncritically may present significant challenges for the clinician who develops and implements treatment plans designed to target each of the six components of the Psychological Flexibility Model (McEnteggart et al., 2015; Muñoz-Martínez & Follette, 2019). More specifically, if the relationship between middle-level terms and basic behavioral principles which supposedly underpin them is not well-described, these concepts may lose utility or even efficacy in the therapeutic setting (Assaz et al., 2018; McEnteggart et al., 2015). First, the term could stray from the philosophical and theoretical foundation of behavior science, approaching the conceptual status of folk psychology. Second, it may decrease the

[1] The ACT model has also been presented in other manners, but the Psychological Flexibility Model is still its mostly disseminated and recognizable form.
[2] Precision means that the concept is so clearly defined and unambiguously differentiated from others that it evokes relatively uniform and constant responses from the members of the verbal community (Biglan & Hayes, 2016).

likelihood of the term being integrated into a functional-analytic and process-based intervention (Callaghan & Darrow, 2015; Hayes et al., 2019), potentially leaving clinicians stranded without adequate knowledge of when and how to use a particular intervention effectively and efficiently.

Fostering the conceptual and empirically-tested link between middle-level terms and basic behavioral principles should be sought to enhance the clinician's understanding and application of ACT-based interventions. Indeed, such a priority is advocated by the CBS "manifesto" (Hayes et al., 2012) and reiterated in the third of a list of recommendations recently made by a CBS task-force (Hayes et al., 2021).

Many researchers within CBS often think of ACT methods as having grown out of Relational Frame Theory (RFT; Hayes et al., 2001), a modern behavior-analytic account of language and cognition which has co-evolved with the therapeutic model. A cursory glance at the literature provides numerous examples of authors advocating that ACT's relevant processes are understood technically by RFT. For example, stating that "Core ACT processes (e.g., defusion, self-as-context) are derived from RFT principles…" (Levin et al., 2020, p. 4); "Acceptance and Commitment Therapy is a behavior change method based on RFT…" (Zhang et al., 2018, p. 2350); "We have given this extremely brief and nontechnical account of RFT to emphasize that ACT is based on a theory of language and cognition…" (Flaxman et al., 2013, p. 13); and "ACT rests on an underlying theory of human language and cognition called relational frame theory" (Harris, 2009, p. 2). Thus, it is clear that it is a widely held view that ACT's components, such as defusion, are on a solid empirical footing of RFT research.

However, some clinicians have highlighted that while ACT is thought to rest upon the empirical framework of RFT, ACT therapists rarely use RFT terminology or RFT-inspired methods directly with clients. For example, Villatte (2020) noted that:

> While RFT concepts have often been linked to ACT in books and trainings, their direct relevance to clinical interventions has not always been clear...However, ACT therapists generally don't use RFT principles to select and craft acceptance and mindfulness interventions. They know why using acceptance and mindfulness makes sense from an RFT perspective, but not how to use RFT to conduct acceptance and mindfulness interventions (p. 30).

We would go further and suggest that it might be reasonable to question Villatte's assertion that clinicians or therapists know why using ACT-based interventions makes sense from an RFT perspective given that they rarely, if ever, refer to RFT principles or processes in practice. Furthermore, some authors have argued that the link between middle-level terms and basic behavioral process or RFT principles seems to be poorly elaborated in ACT, including the central component of cognitive defusion (Assaz et al., 2018; McEnteggart et al., 2015).

This matter is of importance to the clinician because cognitive defusion is one of ACT's most well-researched and defining components in comparison to other therapeutic approaches (e.g., cognitive restructuring in Cognitive-Behavior Therapy; see Hayes, 2022). In other words, cognitive defusion is a signature component that marks ACT out as a theoretically-distinct form of psychotherapy and a unique model of behavior change. As such, it is deserving of closer theoretical and conceptual scrutiny to facilitate the enhancement of therapeutic skill and practice.

The present paper's objective is to increase the pragmatic utility of the concept of "cognitive defusion" by providing a more nuanced, multi-faceted and process-based definition of the term. This will be done through conceptual analysis, highlighting contradictions, fallacies, ambiguities and limitations of the relevant literature (Laurenti et al., 2016; Machado et al., 2000). In the next section, RFT's main concepts will be

briefly presented after which the surveyed ACT literature regarding defusion will be critically examined by means of the conceptual analysis strategies outlined in Machado and Silva (2007). This will culminate in a conceptualization of defusion in terms of its relationship to basic behavioral concepts, that will allow for greater fidelity between therapeutic interventions and expected outcomes for clients.

**Relational Frame Theory**

Sidman and colleagues' groundbreaking work on the emergence of derived equivalence relations among stimuli (Sidman, 1994) sparked an enormous interest in derived relational responding. In essence, these studies demonstrated that if participants learned to respond to arbitrarily defined (i.e. symbolic) relations between stimuli, they could derive relational responses that were not directly taught (Sidman, 1994). These derived relations were named symmetry (if A-B, then B-A) and transitivity (if A-B and B-C, then A-C). Furthermore, subsequent studies showed that relevant stimulus functions (e.g., naming responses) can transfer from one stimulus to another in accordance with the derived relations between the stimuli. For example, if two stimuli, A and B are in a derived equivalence relation with each other, and A is established as a conditioned stimulus for a fear response, this response will spontaneously emerge for B (cf. Dougher et al., 1994).

Relational Frame Theory is a behavior-analytic scientific endeavor to functionally explain the phenomenon of derived relational responding and extend it to interpretations and applications of complex human behavior, especially those related to language and cognition (Hayes et al., 2001; Hughes & Barnes-Holmes, 2016). First, RFT entertains the idea that a variety of relations among stimuli besides equivalence, occurs in natural language, such as opposition, distinction, comparison, hierarchy, causality, temporality, spatiality and deictic (perspective-based). Each of these varied

responding patterns are termed relational frames (Hayes et al., 2001), and each is associated with different pattern of transfer (or more correctly transformation; see Dymond & Rehfeldt, 2000) of functions. For example, if two stimuli, A and B are in a derived comparative relation with each other, such that A is more than B, and B is established as a conditioned stimulus for a fear response, this response will spontaneously emerge, but also amplify for A (see Dougher et al., 1997).

Based on this variety of derived relational responding, RFT's fundamental argument is that to respond to a stimulus by means of arbitrarily defined relations with other stimuli is a generalized operant behavior called Arbitrarily Applicable Relational Responding, or AARR (Hayes et al., 2001). The defining features of AARR are mutual entailment, combinatorial entailment and transformation of function, broader concepts that those of symmetry, transitivity and transfer of function, respectively, more suited to non-equivalence relations (e.g., if A is bigger than B, B is smaller than A).

AARR is approached by RFT as a contextually controlled operant behavior, influenced by antecedent and consequential operations and acquired through a history of multiple exemplar training (Hughes & Barnes-Holmes, 2016), which normally takes place in the social environment (Critchfield & Rehfeldt, 2019). In this learning process, this behavior is brought under the control of two types of contextual cues (Hayes et al., 2001). Relational Contexts (Crel) function to signal the specific type of relation established between stimuli. In a natural social environment, words such as "similar", "like" or "same" may function as a Crel for equivalence relations, evoking responding accordingly. Likewise, "different" or "not similar" may operate as Crels for distinction relations between stimuli. For example, the word *sofa* may be responded to as the same as the word *chair* and different from *oven*.

Furthermore, Functional Contexts (Cfunc) exert control over the type of function transformed through AARRing. More specifically, two stimuli in a coordination frame may be similar in terms of a multitude of stimulus properties. Cfunc stimuli control which of these the stimulus functions are transformed in any given instance. For example, sofa may be similar to chair in the context of *places to sit*, but similar to closet in terms of *price* (i.e., chair is usually *less* expensive).

RFT has accumulated empirical evidence in favor of its account of AARR as a generalized operant behavior, experimentally demonstrating that people can indeed *learn* to respond to arbitrary stimuli through multiple stimulus relations (e.g., Dymond & May, 2018; McLoughlin et al., 2020). Its concepts considerably ramp up the explanatory power of behavioral accounts of complex behavior and have been used to provide a behavioral account of the emergence of responding patterns of clinical interest and that appear to have developed in the absence of any identifiable history of reinforcement in the client's social-cultural environment (Dymond et al., 2018; Guinther & Dougher, 2015), including the development of a whole host of psychological skill sets (Dymond & Rehfeldt, 2000; McLoughlin et al., 2020) and psychological suffering (Dymond et al., 2018). More specifically, it is now within our empirical remit to explain how any stimulus can evoke or control a very large range of responses dependent on the individual's particular history of AARR and the relational (Cfunc and Crel) control being exercised at a particular time.

Over time, it is supposed that a verbally-able human responds to increasingly larger portions of the environment in terms of verbally acquired stimulus functions. Therefore, it is not unusual for a person to respond to an event mostly or entirely based on derived stimulus functions, despite other non-arbitrary stimulus properties being present (e.g., sensory or simple discriminative stimulus properties). ACT therapists

refer to this phenomenon as one of “literality”, and somewhat more technically use the middle-level term “cognitive fusion” to describe it (Hayes et al., 2011).

Cognitive fusion is not necessarily detrimental to the individual. Indeed, it is potentially helpful when the person is responding to contingencies that may be novel, complex, dangerous or that will yield long-term consequences (Hayes et al., 2011). However, depending on the particular responses that are evoked by function transformation and how sensitive it is to the current context, it can cause psychological suffering. Specifically, an evoked response may compete with alternative ones that would yield a higher rate or magnitude of reinforcement in the long-term or that are more congruent with the personal values of the individual. In addition, a high probability of responding to a stimulus in terms of a particular verbal relation can narrow the behavioral repertoire and reduce learning opportunities. In effect, learning becomes less sensitive to other sources of potential behavioral control (i.e., variability).

In a corollary conceptual development, ACT also provided a therapeutic strategy known as *cognitive defusion*, which is intended to disrupt verbal processes that support the transformation of stimulus functions (Assaz et al., 2018; Blackledge, 2007).

**A place for everything and everything in its place**

Within the ACT model, when cognitive fusion is a problem, cognitive defusion is the proposed alternative solution (Hayes et al., 2011; Luoma et al., 2017). But what *precisely* is defusion? Is it a procedure, a process or a state? In order to answer this question, it is worth clarifying that in the behavior-analytic tradition behavior is understood as a continuous flow of organism-environment interactions: an act-in-context. In addition, given certain environmental conditions, the organism has a tendency (i.e. high probability) to act in a specific manner: this enhanced probability represents a “*state*” (cf. Lopes, 2008; Ryle, 1949). At any moment, environmental

variables can be manipulated, and this manipulation represent a *procedure*. This procedure may alter the functional dynamic between organism and environment over time, an effect that we call a process (Catania, 2007)[3]. At the end of the process, a new state, distinct from the previous, is achieved (also referred to as outcome).

One of the main conceptual problems with defusion is that this term is used in the ACT literature to describe all three of the foregoing distinct phenomena: procedure, process and outcome (Assaz et al., 2018; Barnes-Holmes et al., 2016; McEnteggart et al., 2015). At times, defusion is spoken of as something that the therapist does (a procedure): "Defusion techniques used in ACT include meditative exercises, experiential exercises, metaphor, and language conventions" (Luoma et al., 2017, p. 90). Sometimes, it is spoken of as something that is altered in the organism-environment relation (a process): "Cognitive defusion refers to the process of reducing the automatic emotional and behavioral functions of thoughts by increasing awareness of the process of thinking over and above the content or literal meaning of thought" (Blackledge, 2018, p. 351). At other times, it is referred to as something that is achieved (an outcome): "In a state of defusion, you can see a thought for what it is: nothing more or less than a bunch of words or pictures 'inside your head'" (Harris, 2009, p. 21).

This disparity in term usage is not simply the result of distinct authors disagreeing with one another, because it is often found within the same text. On some occasions, these three definitions are so interchangeable that they may all appear in the same paragraph without arising suspicion from a less attentive reader, as follows:

[3] The modern literature on clinical behavior analysis is fairly confusing on what constitutes a process. It has been referred to as any psychological change that occurs without reference to its causes (Barnes-Holmes et al., 2020); changes that occur in predictable sequences towards an outcome (Hayes et al., 2019); and, more broadly, as the therapist-client interaction and subsequent changes over time (Orlinsky et al., 2015). The definition of process adopted in this paper more closely resembles that of mechanism, that explains how a procedure promotes changes (Kazdin, 2007), and presented in behavior analytic textbooks (e.g. Catania, 2007).

> Defusion is a term coined by Hayes and colleagues (1999) describing the attempt to reduce cognitive fusion, or the impact of the transformation of stimulus functions when a client is presented with a verbal event [outcome]. The defusion process aims to disconnect (de-fuse) the person's behavior from the stimulus control of the words [process]. Defusion practices are useful in their own right, and are often integrated into acceptance, contact-with-the-present-moment, and perspective-taking work [procedure]. (Bach & Moran, 2008, p. 166)

To use the same concept to describe a procedure, a process and an outcome is, at the very least, ambiguous and decreases its precision. Considering that precision is one of the stated goal of CBS, this practice presents a problem. At worst, conflating all three phenomena into a single word greatly increases the chances of making category mistakes (cf. Ryle, 1949). Therefore, to increase conceptual clarity, it is best if the tripartite ambiguity of defusion is resolved. As stated by McEnteggart et al. (2015), "we are neither denying that defusion techniques exist, nor that fusion can be reduced, nor that any of this cannot happen through a process of defusion, instead we are simply saying that the same concept cannot be all three types of phenomena" (p. 57).

**Cognitive defusion as an outcome**

As mentioned previously, cognitive fusion can be both beneficial or detrimental to the individual. When it is detrimental and contributing to psychological suffering, it may be specified as a target in case formulation for clinical intervention. One possible intervention, widely used by cognitive-behavioral therapists is cognitive restructuring (Dobson & Dozois, 2010). This procedure is a collaborative endeavor between client and therapist in which both assess and dispute the client's thoughts, appraisals and beliefs (underlying hypothetical constructs in cognitive theory) in terms of their logic

and empirical evidence, with the objective of altering their content (Dobson & Dozois, 2010). An RFT-informed analysis suggests that cognitive restructuring may highlight the incoherence between the client's verbal response and other verbal relations or on-going contingencies, evoking more context-sensitive verbal responses (Blackledge et al., 2009). Thus, verbal control over behavior would persist, but differently than before. This is similar to what clinical behavior analysts do when attempting to change client's self-rules (Hayes et al., 1989).

An alternative to restructuring, defusion arose from the insight that thoughts do not necessarily exert control over behavior, with some contextual conditions being necessary to support this type of functional behavior-behavior relation (Hayes & Brownstein, 1986). From an RFT-perspective, thinking is AARRing. Therefore, its influence over behavior is indirect, by transforming the function of stimuli, thereby evoking responses congruently. This means that, to change the client's verbally controlled responding, it might not be necessary to alter their thoughts' content (the symbolic relations between stimuli), but may be beneficial to target stimulus function transformation. In fact, some authors state that the outcome of defusion is the disruption of function transformation in AARR, reducing the control of thinking over behavior:

> Cognitive defusion techniques attempt to alter the undesirable functions of thoughts and other private events, rather than trying to alter their form, frequency or situational sensitivity. Said another way, ACT attempts to change the way one interacts with or relates to thoughts by creating contexts in which their unhelpful functions are diminished. (Hayes et al., 2006, p. 8).

Similarly, Blackledge (2007) states that:

> …defusion techniques are not intended to change the way people think about their experience. Rather, they are meant to disrupt the uniquely verbal processes

> that give rise to these problematic transformations of function in the first place…In other words, cognitive fusion refers to contexts in which verbal transformations of function are readily occurring, while cognitive defusion refers to contexts in which these verbal transformations are at least temporarily disrupted (p. 557).

Meanwhile, other authors suggest that disrupting function transformation also enables other sources of influence to exert more control over behavior, potentially evoking more useful responses that wield more reinforcement in the long-term. For example, Luoma et al. (2007) argued that: "As a result, the literal functions of problematic thoughts are less likely to dominate as a source of influence over behavior, allowing more helpful, direct, and varied sources of control over action to gain ground" (p. 89). Congruently, "Defusion does not eliminate verbal meaning—it just reduces its automatic effect on behavior such that other sources of behavioral regulation can better participate in the moment" (Hayes et al., 2011, p. 245).

These quotes suggest a precedence (in importance or temporality) of the reduction of verbally acquired stimulus functions over the increase of alternative sources of behavioral influence. However, there are reasons to reconsider this conceptualization. First, to include potential alternative sources of behavioral influence in the definition of cognitive defusion is more coherent with the usual clinical behavior-analytic approach, which aims to broaden repertoires and make them more flexible, rather than focus on the reduction or elimination of problematic responding (Hayes, 2004). Second, a shifted focus towards increasing behavioral variability, rather than reducing sources of behavioral inflexibility, is also more coherent with basic behavioral research, which suggests that a single operation may have multiple effects on behavior. For example, reinforcement (as a procedure) increases the rate of the reinforced

response class while simultaneously decreasing the rate of other response classes and altering the functions of stimuli that acquire discriminative and conditioned reinforcer functions (Catania, 2007).

Given the foregoing, we propose a new understanding of defusion in which it is conceptualized as an outcome characterized by broadened stimulus control over behavior. That is, we propose that defusion has occurred when there is reduced control by previously established verbal stimulus functions and increased control by other sources *without a change in the symbolic relations between the relevant stimuli*. This alteration in stimulus control implies a change in the probability of response classes. That is, in fusion, one particular response is very likely to occur in the presence of a given stimulus (i.e. the "fused" response, congruent with the word stimulus thought's content). In defusion, however, there are multiple competing responses with significant probabilities (Figure 1). One clear implication of this is that fusion and defusion are not all-or-nothing, but should be thought of as a continuum, with higher or lower control of a single verbally acquired function (Blackledge, 2007).

[Insert Figure 1 here]

The conceptualization of defusion as an outcome and not a procedure or a process is made explicit in order to avoid conflating these phenomena into a single term and thus, increase its pragmatic utility. It will be clearer to the reader further in the text, but this conceptual decision is based on the observation that when authors talk about defusion, they refer to a specific (albeit broad) outcome outlined above in contrast with a multitude of procedures and processes (Assaz et al., 2018).

**Procedures to promote cognitive defusion**

A number of different procedures that may be effective in promoting defusion are described throughout the ACT literature (Blackledge, 2015; Hayes et al., 2011;

Hayes & Strosahl, 2004; Luoma & Hayes, 2009). In line with ACT's principles, these tend to be experiential exercises and metaphors. No consensual taxonomy for categorizing these procedures exists. Thus, the one provided below, based on Assaz et al. (2018), is designed to be useful to understand potential interventions and linking them to processes. It is important to point out that these are not fixed boundaries, which means that more complex procedures may be ascribed to more than one category and some of them may appear under the headline of other components of ACT's model (e.g., contact with the present moment or self-as-context) depending on the author.

*Fostering metacognitive awareness* as a strategy includes procedures that require the client to attend not only to their thoughts as discrete occurrences but also to the process of thinking. On the one hand, these may require focus on the stream of thinking, observing and describing their ongoing experience, with imagery instructions to aid (e.g., exercises such as *Leaves on a stream, Soldiers in the parade* or *Watching the mind train*). The client may also be asked to differentiate between different types of thinking patterns, such as evaluations as distinct from descriptions (*Bad cup metaphor*), and inclusive as distinct from oppositional language (using words such as "and" to replace "but" in sentences; *Be out*) and to label recurrent thinking patterns (*Cubbyholing*).

*Playing with words* encompass procedures in which the client is asked to engage in playful activities with a specific thought, such as repeating it out loud (*Word repetition*); singing it to a familiar tune (*Singing thoughts*); speaking it with a peculiar voice, such as that of a cartoon character (*Silly voices*); pronouncing it slowly, emphasizing each syllable (*Slow speech*); or using its synonym in a foreign language (*Word translation*).

*Objectifying thoughts* comprise exercises that attribute physical properties to thoughts, such as an object with size, shape, weight and texture (*Physicalizing*);

passengers on a bus that the client is driving (*Passengers on a bus*); hands that can be moved more far or near from one's face (*Hands as thoughts*); or written words on a sheet of paper (*Content on cards*) or computer screen (*Two computers*). Some of the exercises described above as fostering metacognitive awareness use imagery instructions that overlap with this category.

*Evaluating functional utility* are reflective interventions, in which the therapist enquires about the implications of acting in a manner that is congruent with the thought's content, highlighting potential consequences. These include asking how useful is the thought in pursuing valued-based living; inquiring whether the person or the thoughts are in charge of their actions; or agreeing with the thought's content and asking as to what the client wants to do.

*Disobeying thoughts* comprise procedures designed to create an incongruent experience between what the client is thinking and doing, such as to perform an action after convincing oneself not to do it (*Thoughts and feelings aren't causes*) or writing an order to not do it on a piece of paper placed in their pocket (*Carrying cards*); or an interactive role-play in which therapist and client alternate in the role of the mind, which constantly talks, and the client, who cannot talk but choses what to do (*Take your mind for a walk*). "*Passengers on a bus*" would also fit in this category.

*Distinguishing thought from thinker* includes procedures that actively distinguish between the thoughts (as a psychological experience, a response from the individual) and the person who is having them (the thinker). For example, referring to the mind as a separate entity, that can even be named (*The mind*); thanking the mind for a thought (*Mental appreciation*); including the prefix "I am having a thought that" before describing them (*Having thoughts*); or separating between multiple experiences across time and contexts and the person experiencing them (*Observer*).

*Understanding thinker's function* refers to a set of procedures designed to help the client comprehend how thinking works. This may highlight that our thoughts are arbitrarily defined relations, influenced by our previous history (*Mary had a little...*), that could be easily constructed within a couple of seconds (*What are the numbers*) and altered if needed (*Creating a new story*). Or they may highlight that thoughts are responses made by the individual and, thus, are actions controlled by their effects, whether it be maintaining a narrative coherence (*Master storyteller*) or "selling an idea" increasing the probability of particular actions, with consequences (*Buying thoughts*).

**Processes responsible for cognitive defusion: Altering the functional context**

One of the first attempts to provide a comprehensive and RFT-based account of the process that promote defusion was offered by Blackledge (2007). Based on existing definitions of defusion and RFT concepts, he argued that:

> From an RFT perspective, verbal transformations of stimulus functions are contextually controlled (...) In other words, certain contextual conditions must be in place for verbally specific processes to change stimulus functions (i.e., for cognitive fusion to occur). *We can thus logically assume* [emphasis added] that changing these contextual conditions in certain ways would lead to a disruption of these verbally based functional transformations. This, it is argued, is the essence of cognitive defusion. (Blackledge, 2007, p. 561).

One of the strengths of this definition is that it is informed by basic research on RFT that demonstrates contextual control over function transformation (e.g., Dougher et al., 2002; Perez et al., 2015; Roche et al., 2000). Thus, if contextual cues control function transformation in symbolic relations (i.e. functional context), defusion would involve altering the functions of these same cues. This was an idea that several researchers have since adopted (e.g., Hayes et al., 2011).

However, a thorough conceptual analysis of this idea reveals unacknowledged assumptions or steps in the argument (Machado & Silva, 2007). This definition is based on an implicit logical assumption: it assumes that because fusion and defusion are opposite in terms of states, they should be the opposite in terms of processes (Assaz et al., 2018). This is a "symmetry assumption" that does not necessarily follow and is required to be proven empirically.

A limitation of this assumption (cf. Machado & Silva, 2007) is that it is not always supported by the outcomes of published research. More specifically, symbolically acquired stimulus functions can be reduced through direct learning processes, namely respondent and operant extinction (Bennet et al., 2020; Boldrin & Debert, 2020; Roche et al., 2008; Vervoort et al., 2014), that differ from the procedures that were used to generate the acquired stimulus functions in the first instance. In other words, acquisition and reduction of stimulus function may not work through similar processes. Thus, even if clinical problems do arise by virtue of derived transformations of stimulus functions under tight contextual control (i.e., fusion), it does not follow that the problem is best treated with relational interventions (Kanter, 2013). And the reverse also applies; even where altering the contextual control over the transformation of functions is effective in reducing problematic behavior, it does not necessarily follow that the problematic behavior arose through a derived transformation of response functions. In an analogous way, scurvy is a disease caused by a lack of vitamin C, but it cannot be rectified by the intake of vitamin C.

Another unacknowledged assumption in Blackledge's (2007) definition is that there is a single process by which defusion is achieved: changing the contextual cues that control function transformation. The "single process assumption" is questioned by equifinality studies indicate that similar outcomes may be achieved through different

pathways: a stimulus may acquire similar function and magnitude through distinct learning processes, such as operant conditioning, vicarious learning, instructions and derived relational responding (Cameron et al., 2016; Dymond et al., 2012). In other words, a single outcome may be achieved through different processes.

**Processes responsible for cognitive defusion: Multiple pathways**

*Selective Attending Pathway*

Rejecting the symmetry and single process assumptions for the reasons outlined above, Assaz et al. (2018) proposed an alternative conceptualization. The authors argued that there were multiple process pathways by which the outcome of defusion can be achieved, each of them exemplified by a list of procedures found in the ACT literature (Assaz et al., 2018). However, we have identified some limitations of that account (cf. Machado & Silva, 2007), that can be addressed in order to provide a more nuanced and empirically-based interpretation of defusion.

The most relevant of these is in the argument that *playing with words* exercises are exposure-based procedures because they involve repeated contact with a target thought. One of these interventions, called *Word Repetition*, is among the most widely used procedures in ACT analogue research. Two studies, that manipulated relevant parameters of word repetition (duration and rate), offer valuable insight. Specifically, Masuda et al. (2009) contrasted 1, 3, 10, 20, and 30 seconds word repetition over two experiments and found that the range between 20 and 30 seconds lead to the greater changes in a thought's associated distress and believability. This is consistent with an exposure account: the highest duration yields the larger effect. However, Tyndall et al. (2017) compared the rate of repetition for 30 seconds: 0.5, 1 and 2 words per second. The exposure hypothesis would predict that the third group, given more exposure to the

word, would experience larger effects. But, the one word per second group in fact experienced the greatest reductions in thought discomfort and believability.

To better understand this result, reports of the participants' experience would be helpful. Unfortunately, analog studies rarely provide qualitative data. Luckily, *Word Repetition* is also commonly used in ACT manuals to demonstrate interventions to promote defusion. This enables us to gather anecdotal evidence of common effects of this type of practice, highlighting not only how the symbolically acquired functions (i.e., word meaning) reduced for the clients after repeating it out loud for some time, but also how the formal properties of the word (i.e., the sound) and proprioceptive stimulation (jaw, mouth and tongue movements) become more pronounced (Blackledge, 2015, p. 93-97; Hayes et al., 2011, p. 248-250; Luoma et al., 2017, p. 89).

In line with our definition of defusion as an outcome of broadened stimulus control, with reduced influence of previously verbally acquired functions without altering established symbolic relations between stimuli, repeating a word out loud can shift stimulus control from the word's meaning to its formal properties, whether phonetic or proprioceptive. Similar word play exercises may do the same by drawing attention to how these words or phrases sound, are written, and how we use our bodies to speak them, rather than their verbally attributed meaning. In everyday life, this experience occurs in tongue twisters that highlight phonetic properties of sentences with absurd meanings for humorous effect or in attempting to repeat a word in a foreign language (i.e. echoic behavior; cf. Skinner, 1957).

If that is the case, rather than mere exposure, *playing with words* exercises would work through the process of selective attending to particular stimulus properties. To understand this, it is important to note that all stimuli have multiple potential aspects to which a person might attend to. Preliminary studies on selective attending

demonstrated how specific properties of a stimulus may exert strong control over responding at the expense of others (Lovaas et al., 1979; Ray, 1969; Reynolds, 1961). Later, researchers found that formal properties of the stimuli, such as its location, may exert control over responding to the point of inhibiting symbolic learning (Iversen et al., 1986; Lionello & Urcuioli, 1998). More recently, studies demonstrated that formal properties such as the color of a written word or the speaker's voice are capable of interfering and disrupting already established symbolic functions (Kenny et al., 2014; Stewart et al., 2002; Stewart & Lavelle, 2013). Taken together, these experiments show how formal properties may exert strong stimulus control to the expense of symbolically acquired functions. Based on this, we argue that *playing with words* procedures foster selective attending to the formal (auditory, visual or proprioceptive) properties of stimuli that have verbally acquired functions, altering stimulus control and promoting defusion.

To illustrate, imagine a person who has the recurrent thought that if they were to shake hands with someone, they would get sick. This thinking is, in essence, relational framing, in which the there is an arbitrary relation established between the stimuli "object" and "disease", controlled by a relational context of causality (if, then) and a functional context of danger. Thus, the aversive function of disease may transform the object's function, evoking "fused" responses that are congruent with the thought's content: avoiding contact with the object. The thought as well as the fused response are highlighted in gray at the center of Figure 2. Then, auditory, visual or proprioceptive properties of the stimulus "disease" may be highlighted through *playing with words* exercises and selectively attended to, broadening stimulus control and creating defusion (illustrated on the left portion of Figure 2, outlined as selective attending pathway).

[Insert Figure 2]

*Spatial Distancing Pathway*

The procedures described as *objectifying thoughts* also enhance the salience of physical properties, albeit those arbitrarily attributed to them by the client or the therapist through coordination frames with real objects. Thus, it is likely that selective attention may also partially account for the effects of these procedures.

We emphasize "partially" because verbally characterizing a thought as an object not only highlights its physical properties, but, also, establishes it as something distinct from the individual, that is spatially distant. Thus, spatial and deictic framing might be relevant aspects of these procedures. The experimental literature on spatial framing is incipient (May et al., 2017), while deictic framing has a more extensive history of experimental demonstration, assessment and training (Montoya-Rodriguez et al., 2017). Nevertheless, the RFT literature has not investigated the idea that framing a stimulus as spatially distant from oneself will change its functions.

This research gap is somewhat filled from two lines of investigation. First, a growing body of evidence supports the notion of spatial discounting. Similar to other discounting phenomena, such as delay or probability, this means that there is a tendency for stimuli to exert less control over responding the more distant in space they are from the individual (Mallpress, 2021; Robinson et al., 2019; Stevens et al., 2005). Second, in applied settings, the line of research on self-distancing, has shown in multiple experiments that experiencing a situation or a memory in the third person (distanced perspective) decreases emotional reactivity and impulsivity in comparison to a first person point of view (immersed perspective; Kross & Ayduk, 2017).

Both of these lines of research converge in suggesting that verbally establishing a stimulus as spatially distant reduces its control over behavior. This intuitively makes sense. We tend to respond differently to a dangerous situation if it is occurring in our

immediate perimeter, 10 kilometers away or on the other side of the world. Therefore, objectifying thoughts may highlight the physical properties attributed to them, selectively attended by the client, and simultaneously reduce the control they exert by establishing them as spatially distant from oneself with spatial and deictic AARRing.

Going back to Figure 2, this pathway is illustrated in the bottom part, below the thought. In *objectifying thoughts* procedures, the thought "If I touch their hand, then I will get sick" becomes framed as coordinated with a physical object. This enables two things. First, the person may selectively attend to this object's formal properties, similarly to what is described in the previous section and illustrated on the left of Figure 2. Second, the person can easily derive that the object (which is coordinated with their thoughts) is spatially distant from themselves through deictic and spatial framing, thus reducing its control over behavior.

*Relational Coherence Pathway*

Another potential pathway to achieve cognitive defusion is represented by the procedures categorized as *distinguishing thought from thinker* and *understanding thinker's function*. Taken together, these exercises establish two sets of verbal relations. The first is between the thought and the person who is having the thought (which will be referred to as the thinker), highlighting the fact that this is a verbal response from an individual. These may be described in terms of deictic framing in addition to distinction ("my thoughts are different from me") or hierarchical framing ("my thoughts are part of me"), thus overlapping with ACT's component of self-as-context (Foody et al., 2013; Gil-Luciano et al., 2017; López-López & Luciano, 2017).

The second set of verbal relations establishes the thinker as unreliable. This is done by highlighting that the thinker's behavior is determined by historical and current contingencies and thus, serves a function in the individual's repertoire. These functions

may be varied, but often revolve around two. First, thinking can aid the individual to solve problems (cf. Cassidy et al., 2010), increasing their likelihood of obtaining high magnitude positive or negative reinforcers.

A second function of thinking is maintaining relational coherence with past learning experiences (Bordieri et al., 2016; Quiñones & Hayes, 2014). Relational coherence describes the extent to which a specific pattern of AARRing functionally overlaps with previously reinforced patterns (Barnes-Holmes et al., 2017). According to RFT, relational coherence often functions as a conditioned reinforcer of due to its importance in social communication and problem-solving (Bordieri et al., 2016; Quiñones & Hayes, 2014). If relations are not consistent with each other, the person will have difficulties in responding effectively to their environment. In an absurd scenario, if "cheese" were a tool as well as a type of food, how could one respond effectively to the question "do you want cheese?".

The thinker's behavior, although functionally valid and relevant to solve problems and maintain coherence, is not intended to "represent" reality, but rather help the individual to interact more effectively with their surrounding environment. But, it may do so at a high cost, evoking behaviors to the expense of the person's subjective well-being or life quality. For example, in Figure 2, the person's worries about contamination may be coherent with past events in which they experienced or witnessed someone struggle with a deadly disease; and evoke avoidance responses, diminishing the probability of getting sick ever occurring again (even if the person was not likely to be contaminated in the first place). However, in the long-term, will probably interfere with their ability to connect with other people or force them to wash their hands constantly, in a time-consuming and potentially health hazardous effort.

When the client relates the two established relations (i.e. thought-thinker and thinker-unreliable), illustrated in the top part of Figure 2, the resulting network will initially be low in coherence. Specifically, if the person acts congruently with the verbal relations established in the thought, they are also acting congruently with the instructions of an unreliable speaker, which has shown that they should not be trusted (cf. Bianchi et al., 2021). Because incoherent networks disrupt communication and problem-solving, perceiving incoherent networks can hinder their transformation of function, reducing the thought's control over behavior (cf. Bianchi et al., 2021).

Thus, by framing thoughts as distinct from or part of a speaker, who is unreliable due to its functional and historical nature, the client and the therapist may create an incoherent relational network. These newly formed arbitrary relations between stimuli lead to broader stimulus control and disruption of potential transformation of function caused by thinking. In other words, cognitive defusion. This is not that different from how people learn to respond to dreams, establishing them as responses that we experience privately, but do not "take them seriously" if they are not helpful: if your partner cheats on you on your dreams, this does not mean they are doing so in reality.

It is also similar to how, in interpersonal contexts, people respond differently to a verbalization depending on the speaker. A person will tend to follow rules or not depending on whether the speaker is seen as reliable, knowledgeable and lacking in ulterior motives (cf. Stapleton, 2020). This has been experimentally demonstrated in studies in which the speaker provides inaccurate rules, thus affecting subsequent rule-following (Bianchi et al., 2021; Martinez-Sanchez & Ribes-Iñesta, 1996; Martinez & Tamayo, 2005), and in investigations into how children differentiate between reliable and non-reliable speakers when learning new information, including vocabulary (Mills,

2013). The unique aspect of some procedures to promote defusion is that they help clients to do the same regarding themselves, self-evaluating as unreliable speakers.

*Differential Reinforcement Pathway*

Lastly, the group of exercises categorized as *evaluating functional utility* and *disobeying thoughts* create situations in which, given a specific thought that usually evokes responding congruent with its content (i.e. fused response), the client is encouraged to emit alternative responses to the usual one (i.e. defused response), thereby "disobeying" their thoughts. These are then differentially reinforced, either naturally, by contacting high-magnitude reinforcers or engaging in values-based behaviors; or by the therapist through social reinforcement or by highlighting the natural consequences of disobeying the thought that might be otherwise overlooked by the client. Either way, this could be characterized as a Differential Reinforcement of Alternative Responses (DRA) procedure, in which the usual response is extinguished in favor of new, alternative ones that are reinforced (cf. Vollmer & Iwata, 1992).

The foregoing pathway is illustrated on the right in Figure 2, in which approaching the threatening hand is differentially reinforced as an alternative to avoiding it. In effect, stimulus control is broadened, allowing one increased access to high-magnitude reinforcers if they were to "disobey" the thoughts they are fused.

Two lines of empirical investigation have demonstrated the foregoing processes. The first involves the delivery of reinforcement for new responses that compete with previously learned responses acquired through symbolic relations (e.g., Bennett et al., 2020; Bones et al., 2001; Broothaerts, 2015). The second involves attempting to reduce avoidance response rates that have been established through AARR using exposure methods targeting *either* the original directly conditioned discriminative stimulus for avoidance *or* the derived relational stimuli whose avoidance functions have been

acquired through the AARR process, and comparing outcomes (Boldrin & Debert, 2020; Roche et al., 2008). The results of such studies vary and are hard to interpret collectively, but it is clear at this point that they derived extinction of fear and avoidance functions is possible.

**Conceptualization evaluation and future research directions**

In summary, we argue that cognitive defusion is an outcome of broadened stimulus control over behavior, that has occurred when there is reduced control by previously established verbal stimulus functions and increased control by other sources without altering the symbolic relations between the relevant stimuli. This outcome can be achieved through different processes, each underlying a specific set of procedures found in the ACT literature. We have outlined four potential process pathways: selective attending, spatial distancing, relational coherence and differential reinforcement (Table 1).

[Insert Table 1]

In order to evaluate this conceptualization of defusion, the parameters outlined by Ferreira et al. (2020) for conceptual research in Clinical Behavior Analysis are useful. They propose five criteria to assess a concept's strength: conceptual precision; coherence with underlying philosophy and theory; with basic empirical research; and with applied research; and utility for the practitioner (Ferreira et al., 2020).

To begin with, since our analysis was guided by RFT concepts and behavioral principles and was kept within these boundaries, it is coherent with the underlying theory and their underlying philosophy. Furthermore, by explicitly using the concept of defusion to refer to an outcome and distinguishing it from relevant procedures and process that lead to it, the present conceptualization enhances the concept's precision in comparison to other approaches. It is further strengthened by linking the procedures

and processes responsible for promoting defusion in RFT and basic behavioral concepts, enabling behavior scientists to describe their interventions and the reasons for their effectiveness in more precise terms.

Moreover, the proposed processes relevant to promote cognitive defusion are coherent with behavioral phenomena demonstrated in controlled experimental settings. These include the demonstration of formal properties of stimuli hindering function transformation in AARRing (selective attending pathway); spatial framing reducing stimulus control (spatial distancing pathway); relational incoherence disrupting transformation of function (relational coherence pathway); and reinforcement of alternative behaviors to verbally learned responses (differential reinforcement pathway). Because some of these lines of research are still incipient, further empirical evidence will enable us to corroborate or refine the present conceptualization of defusion.

Regarding applied research, unfortunately there is a gap in the literature relating to research that could be useful to corroborate or refine the present conceptualization. What is needed, are more experimental analog studies that go beyond merely evaluating the effects of specific exercises designed to create defusion. Such research could include manipulating distinct parameters within the same procedure to promote defusion, such as execution parameters (Masuda et al., 2009; Tyndall et al., 2017), types of relational framing (Foody et al., 2013; Gil-Luciano et al., 2017; López-López & Luciano, 2017) and experiential training (Mandavia et al., 2015). Future research could also contrast the efficacy of different procedures or assess relevant moderators to the effectiveness of these procedures, for example.

Lastly, by emphasizing the potential processes responsible for defusion, we hope to help increase the quality of mental health care provided by ACT therapists and provide a much-needed guide in their drive to implement evidence-based practice with

their clients. As Lilienfeld et al. (2013) noted, one of the core components in the implementation of evidence-based practice in psychological therapies, is the capacity of the clinician to identify and employ the strongest available evidence for a treatment, that allows for the choice of the therapeutic intervention deemed to be the most effective for a client presentation. In addition, the therapist should have a clear comprehension of why that treatment is effective. Indeed, it is often understated how much value clinicians place in understanding underlying theory. To illustrate, Huisman and Kangas (2018) reported that for most clinical psychologists, theory and evidence-based practice are key in structuring their case formulations with clients. Indeed, Pilecki and McKay (2013) acknowledged that one of the reasons why some clinical psychologist might not embrace evidence-based practice is that they lack the requisite familiarity with the psychological theory that underpins the treatments.

In the context of the current paper, understanding the functional pathways to defusion allows clinicians to focus on the main mechanisms of change while implementing their interventions; create new exercises that work through the same processes; and adapt procedures to their style and the client's needs and idiosyncrasies. In sum, a process-based analysis of defusion will increase the quality of the therapist's decision-making, enhance functional coherence, creativity and flexibility, and ultimately enhance their effectiveness as practitioners.

We hope to have illustrated how far we have come in developing a well understood technically grounded account of defusion, but also how far we have to go. The challenges ahead are both empirical and conceptual and it is an effort that could fairly be considered to be in the very early stages, despite a considerable body of research in each domain already having been produced. Nevertheless, progress is being made and we hope that the current paper offers framework within which other

researchers can conduct additional conceptual analyses, and of course the complementary empirical analyses that are required to test and hone those ideas. To this extent, we hope that even in its current embryonic form, our analysis will go some way towards helping clinicians make more functionally-analytic informed decisions regarding their therapeutic strategies.

**References**


Assaz, D. A., Roche, B., Kanter, J. W., & Oshiro, C. K. (2018). Cognitive defusion in acceptance and commitment therapy: What are the basic processes of change?. *The Psychological Record, 68*(4), 405-418. https://doi.org/10.1007/s40732-017-0254-z

Bach, P. A.; & Moran, D. J. (2008). *ACT in practice: Case conceptualization in acceptance & commitment therapy*. New Harbinger Publications.

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*(4), 434-445. https://doi.org/10.1016/j.jcbs.2017.08.001

Barnes-Holmes, Y., Barnes-Holmes, D., McEnteggart, C., Dougher, M. J., & Luciano, C. (2020). Introduction to the special issue on clinical psychology and behavioral science: Processes, principles, and analytic strategies. *The Psychological Record, 70,* 541-542. https://doi.org/10.1007/s40732-020-00444-9

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. 365-382). John Wiley & Sons.

Bennett, M. P., Roche, B., Dymond, S., Baeyens, F., Vervliet, B., & Hermans, D. (2020). Transitions from avoidance: Reinforcing competing behaviours reduces generalised avoidance in new contexts. *Quarterly Journal of Experimental Psychology, 73,* 2119-2131. 10.1177/1747021820943148

Bianchi, P. H., Perez, W. F., Harte, C., & Barnes-Holmes, D. (2021). Effects of coherence on speaker preference and rule-following. *Perspectivas em Análise do Comportamento, 12*(1). https://doi.org/10.18761/PAC.2021.v12.RFT.07

Biglan, A., & Hayes, S. C. (2016). Functional contextualism and contextual behavioral science. In: R.D. Zettle, S.C. Hayes, D. Barnes-Holmes & A. Biglan (Eds.). *The Wiley Handbook of Contextual Behavioral Science* (pp. 37-61). John Wiley & Sons.

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

Blackledge, J. T. (2015). *Cognitive defusion in practice: A clinician's guide to assessing, observing & supporting change in your client*. Context Press.

Blackledge, J. T. (2018). Cognitive defusion. In: S. C. Hayes & S. G. Hoffman (Eds.) *Process-based CBT: The science and core clinical competencies of cognitive-behavioral therapy (pp. 351-361).* Context Press.

Blackledge, J. T., Moran, D. J., & Ellis, A. (2009). Bridging the divide: Linking basic science to applied psychotherapeutic interventions - a relational frame theory account of cognitive disputation in rational emotive behavior therapy. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *27*(4), 232-248. https://doi.org/10.1007/s10942-007-0078-x

Boldrin, L. S., & Debert, P. (2020). Direct and derived extinction of avoidance responses in equivalence classes. *The Psychological Record*, *70*(3), 433-444. https://doi.org/10.1007/s40732-020-00419-w

Bones, R., Keenan, M., Askin, D., Adams, L., Taylor, D., & Nicholas, O. (2001). The effects of response topography on functional equivalence class formation. *The Psychological Record*, *51*(1), 89-110. https://doi.org/10.1007/BF03395388

Bordieri, M. J., Kellum, K. K., Wilson, K. G., & Whiteman, K. C. (2016). Basic properties of coherence: Testing a core assumption of relational frame theory. *The Psychological Record*, *66*(1), 83-98. https://doi.org/10.1007/s40732-015-0154-z

Broothaerts, K. (2015). *Acceptance and commitment therapy: A scientific assessment of the cognitive defusion technique* [Unpublished master's thesis]. Katholieke Universiteit Leuven.

Callaghan, G. M., & Darrow, S. M. (2015). The role of functional assessment in third wave behavioral interventions: Foundations and future directions for a fourth wave. *Current Opinion in Psychology*, *2*, 60-64. https://doi.org/10.1016/j.copsyc.2014.12.005

Cameron, G., Roche, B., Schlund, M. W., & Dymond, S. (2016). Learned, instructed and observed pathways to fear and avoidance. *Journal of Behavior Therapy and Experimental Psychiatry*, *50*, 106-112. https://doi.org/10.1016/j.jbtep.2015.06.003

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational frame theory and human intelligence. *European Journal of Behavior Analysis*, *11*(1), 37-51. https://doi.org/10.1080/15021149.2010.11434333

Catania, A. C. (2007). *Learning (4th edition)*. Sloan Publishing.

Critchfield, T. S., & Rehfeldt, R. A. (2019). Engineering emergent learning with nonequivalence relations. In: J. O. Cooper, T. E. Heron, & W. L. Herald *Applied behavior analysis (3rd edition) (pp.* 497-526). Pearson

Dobson K. S., & Dozois, D. J. (2010) Historical and philosophical bases of the cognitive-behavioral therapies. In: Dobson K.S. (Ed.). *Handbook of cognitive-behavioral therapies* (3rd edition) (pp. 3-38). Guilford Press.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *62*(3), 331-351. https://doi.org/10.1901/jeab.1994.62-331

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior*, *88*(2), 179-197. https://doi.org/10.1901/jeab.2007.45-05

Dougher, M., Perkins, D. R., Greenway, D., Koons, A., & Chiasson, C. (2002). Contextual control of equivalence-based transformation of functions. *Journal of the Experimental Analysis of Behavior*, *78*(1), 63-93. https://doi.org/10.1901/jeab.2002.78-63

Dymond, S., & May, R. (2018). Quantifying the empirical growth of Relational Frame Theory research: A cautionary note. *The Psychological Record*, *68*(2), 255-260. https://doi.org/10.1007/s40732-018-0278-z

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, *23*(2), 239-254. https://doi.org/10.1007/BF03392013

Dymond, S., Bennett, M., Boyle, S., Roche, B., & Schlund, M. (2018). Related to anxiety: Arbitrarily applicable relational responding and experimental psychopathology research on fear and avoidance. *Perspectives on Behavior Science*, *41*(1), 189-213. https://doi.org/10.1007/s40614-017-0133-6

Dymond, S., Schlund, M. W., Roche, B., De Houwer, J., & Freegard, G. P. (2012). Safe from harm: learned, instructed, and symbolic generalization pathways of human threat-avoidance. *PLoS One*, *7*(10), e47539. https://doi.org/10.1371/journal.pone.0047539

Ferreira, T. A. D. S., Simões, A. S., Santos, F. M., Matos, J. P., & Moura, M. C. (2020). Methods of Conceptual Research in Clinical Behavior Analysis. *The Psychological Record*, 1-8. https://doi.org/10.1007/s40732-020-00411-4

Flaxman, P. E., Bond, F. W., & Livheim, F. (2013). *The mindful and effective employee: An acceptance and commitment therapy training manual for improving well-being and performance*. New Harbinger Publications.

Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., & Luciano, C. (2013). An empirical investigation of hierarchical versus distinction relations in a self-based ACT exercise. *International Journal of Psychology and Psychological Therapy, 13*(3), 373–388.

Gil-Luciano, B., Ruiz, F. J., Valdivia-Salas, S., & Suárez-Falcón, J. C. (2017). Promoting psychological flexibility on tolerance tasks: Framing behavior through deictic/hierarchical relations and specifying augmental functions. *The Psychological Record*, *67*(1), 1–9. https://doi.org/10.1007/s40732-016-0200-5

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 Behavioral Science*, *18*, 181-192. https://doi.org/10.1016/j.jcbs.2020.09.009

Guinther, P. M., & Dougher, M. J. (2015). The clinical relevance of stimulus equivalence and relational frame theory in influencing the behavior of verbally competent adults. *Current Opinion in Psychology*, *2*, 21-25. https://doi.org/10.1016/j.copsyc.2015.01.015

Harris, R. (2009). *ACT made simple: An easy-to-read primer on acceptance and commitment therapy.* New Harbinger Publications.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy*, *35*(4), 639-665. https://doi.org/10.1016/S0005-7894(04)80013-3

Hayes, S. C. (2022). Defusion and acceptance as treatment processes in cognitive behavioral therapy. *Cognitive and Behavioral Practice*. https://doi.org/10.1016/j.cbpra.2022.01.005

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst*, *9*(2), 175-190. https://doi.org/10.1007/BF03391944

Hayes, S. C., & Strosahl, K. D. (Eds.) (2004). *A practical guide to acceptance and commitment therapy*. Springer.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* Kluwer Academic.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human

condition. *Journal of Contextual Behavioral Science, 1*(1-2), 1-16. https://doi.org/10.1016/j.jcbs.2012.09.004

Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy, 117*, 40-53. https://doi.org/10.1016/j.brat.2018.10.005

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1-25. https://doi.org/10.1016/j.brat.2005.06.006

Hayes, S. C., Merwin, R. M., McHugh, L., Sandoz, E. K., A-Tjak, J. G., Ruiz, F. J., Barnes-Holmes, D., Bricker, J. B., Ciarrochi, J., Dixon, M. R., Fung, K. P., Gloster, A. T., Gobin, R. L., Gould, E. R., Hofmann, S. G., Kaujja, R., Karekla, M., Luciano, C. & McCracken, L. M. (2021). Report of the ACBS task force on the strategies and tactics of contextual behavioral science research. *Journal of Contextual Behavioral Science, 20*, 172-183. https://doi.org/10.1016/j.jcbs.2021.03.007

Hayes, S. C., Wilson, K. G., & Strosahl, K. D. (2011). *Acceptance and commitment therapy: The process and practice of mindful change (2nd version).* The Guilford Press.

Hayes, S. C., Zettle, R. D., & Rosenfarb, I. (1989). Rule-following. In: S. C. Hayes (Ed.), *Rule-Governed Behavior: Cognition, contingencies and instructional control* (pp. 191-220). Springer.

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

Huisman, P., & Kangas, M. (2018). Evidence-based practices in cognitive behaviour therapy (CBT) case formulation: What do practitioners believe is important, and what do they do?. *Behaviour Change*, *35*(1), 1-21. https://doi.org/10.1017/bec.2018.5

Iversen, I. H., Sidman, M., & Carrigan, P. (1986). Stimulus definition in conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *45*(3), 297-304. https://doi.org/10.1901/jeab.1986.45-297

Kanter, J. W. (2013). The vision of a progressive clinical science to guide clinical practice. *Behavior Therapy*, 44(2), 228-233. https://doi.org/10.1016/j.beth.2010.07.006

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology.*, *3*, 1-27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kenny, N., Barnes-Holmes, D., & Stewart, I. (2014). Competing arbitrary and non-arbitrary relational responding in normally developing children and children diagnosed with autism. *The Psychological Record, 64*(4), 755-768. https://doi.org/10.1007/s40732-014-0042-y

Kross, E., & Ayduk, O. (2017). *Self-distancing: Theory, research, and current directions.* In: J. M. Olson (Ed.), *Advances in experimental social psychology* (p. 81–136). Elsevier Academic Press.

Laurenti, C., Lopes, C. E., & Araújo, S. F. (2016). *Pesquisa teórica em psicologia: Aspectos filosóficos e metodológicos.* Hogrefe CETEPP.

Levin, M., Krafft, J., & Twohig, M. P. (2020). Refining Acceptance and Commitment Therapy to meet the challenge of the human condition. In: M. E. Levin, M. P.

Twohig, and J. Krafft (eds.), I*nnovations in Acceptance and Commitment Therapy: Clinical Advancements and Applications in ACT* (pp. 1-13). Context Press.
Lilienfeld, S., Ritschel, L., Lynn, S.J., Cautin, R., & Latzman, R.D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review, 33,* 883–900. doi:10.1016/j.cpr.2012.09.008
Lionello, K. M., & Urcuioli, P. J. (1998). Control by sample location in pigeons' matching to sample. *Journal of the Experimental Analysis of Behavior*, *70*(3), 235-251. https://doi.org/10.1901/jeab.1998.70-235
Lopes, C. E. (2008). Uma proposta de definição de comportamento no behaviorismo radical. *Revista Brasileira de Terapia Comportamental e Cognitiva, 10*(1), 1-13. https://doi.org/10.31505/rbtcc.v10i1.206
López-López, J. C., & Luciano, C. (2017). An experimental analysis of defusion interactions based on deictic and hierarchical framings and their impact on cognitive performance. *The Psychological Record, 67*(4), 485–497. https://doi.org/10.1007/s40732-017-0250-3
Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86*(6), 1236–1254. https://doi.org/10.1037/0033-2909.86.6.1236
Luoma, J. & Hayes, S. C. (2009). Cognitive defusion. In: W. T. O'Donohue, & J. E. Fisher (Eds.), *Empirically supported techniques of Cognitive Behavior Therapy: A step by step guide for clinicians (2nd edition)* (pp. 181-188). John Wiley & Sons.
Luoma, J. B., Hayes, S. C., & Walser, R. D. (2017). *Learning ACT: An Acceptance & Commitment Therapy skills-training manual for therapists (2nd version)*. New Harbinger Publications.
Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist, 62*(7), 671-681. https://doi.org/10.1037/0003-066X.62.7.671
Machado, A., Lourenço, O., & Silva, F. J. (2000). Facts, concepts, and theories: The shape of psychology's epistemic triangle. *Behavior and Philosophy*, *28*(1-2), 1-40.
Mallpress, D. E. (2021). Some theoretical notes on spatial discounting. *Behavioural Processes, 186,* 104355. https://doi.org/10.1016/j.beproc.2021.104355
Mandavia, A., Masuda, A., Moore, M., Mendoza, H., Donati, M. R., & Cohen, L. L. (2015). The application of a cognitive defusion technique to negative body image thoughts: A preliminary analogue investigation. *Journal of Contextual Behavioral Science, 4*(2), 86-95. https://doi.org/10.1016/j.jcbs.2015.02.003
Martinez, H., & Tamayo, R. (2005). Interactions of contingencies, instructional accuracy, and instructional history in conditional discrimination. *The Psychological Record*, *55*(4), 633-646. https://doi.org/10.1007/BF03395531
Martinez-Sanchez, H., & Ribes-Iñesta, E. (1996). Interactions of contingencies and instructional history on conditional discrimination. *The Psychological Record*, *46*(2), 301-316.
Masuda, A., Hayes, S. C., Twohig, M. P., Drossel C., Lillis, J., & Washio, Y. (2009). A parametric study of cognitive defusion and the believability and discomfort of negative self-referential thoughts, *Behavior Modification, 33,* 250-262. https://doi.org/10.1177/0145445508326259
May, R. J., Stewart, I., Baez, L., Freegard, G., & Dymond, S. (2017). Arbitrarily applicable spatial relational responding. *Journal of the Experimental Analysis of Behavior*, *107*(2), 234-257. https://doi.org/10.1002/jeab.250

McEnteggart, C., Barnes-Holmes, Y., Hussey, I., & Barnes-Holmes, D. (2015). The ties between a basic science of language and cognition and clinical applications. *Current Opinion in Psychology*, *2*, 56-59. https://doi.org/10.1016/j.copsyc.2014.11.017

McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Convergence of multiple fields on a relational reasoning approach to cognition. *Intelligence*, *83*, 101491. https://doi.org/10.1016/j.intell.2020.101491

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404-418. https://doi.org/10.1037/a0029500

Montoya-Rodriguez, M. M., Molina, F. J., & McHugh, L. (2017). A review of relational frame theory research into deictic relational responding. *The Psychological Record*, *67*(4), 569-579. https://doi.org/10.1007/s40732-016-0216-x

Muñoz-Martínez, A. M., & Follette, W. C. (2019). When love is not enough: The case of therapeutic love as a middle-level term in functional analytic psychotherapy. *Behavior Analysis: Research and Practice*, *19*(1), 103-113. http://dx.doi.org/10.1037/bar0000141

Orlinsky, D., Heinonen, E., & Hartmann, A. (2015). Psychotherapy process research. In: J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (2nd edition) (pp. 515-520)*. Elsevier.

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 Behavior*, *103*(3), 511-523. https://doi.org/10.1002/jeab.150

Pilecki, B. & McKay, D. (2013). The theory-practice gap in cognitive-behaviour therapy. *Behaviour Therapy, 44*(4), 541-547. https://doi.org/10.1016/j.beth.2013.03.004

Quiñones, J. L., & Hayes, S. C. (2014). Relational coherence in ambiguous and unambiguous relational networks. *Journal of the Experimental Analysis of Behavior*, *101*(1), 76-93. https://doi.org/10.1002/jeab.67

Ray, B. A. (1969). Selective attention: The effects of combining stimuli which control incompatible behavior. *Journal of the Experimental Analysis of Behavior*, *12*(4), 539-550. https://doi.org/10.1901/jeab.1969.12-539

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, *4*(3), 203-208. https://doi.org/10.1901/jeab.1961.4-203

Robinson, E., Michaelis, K., Thompson, J. C., & Wiener, M. (2019). Temporal and spatial discounting are distinct in humans. *Cognition*, *190*, 212-220. https://doi.org/10.1016/j.cognition.2019.04.030

Roche, B. T., Kanter, J. W., Brown, K. R., Dymond, S., & Fogarty, C. C. (2008). A comparison of "direct" versus "derived" extinction of avoidance responding. *The Psychological Record*, *58*(3), 443-464. https://doi.org/10.1007/BF03395628

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes-Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record*, *50*(2), 267-291. https://doi.org/10.1007/BF03395356

Ryle, G. (1949). *The concept of mind*. University of Chicago Press.

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

Skinner, B. F. (1957). *Verbal behavior*. Prentice-Hall, Inc.

Stapleton, A. (2020). Choosing not to follow rules that will reduce the spread of COVID-19. *Journal of Contextual Behavioral Science*, *17*, 73-78. https://doi.org/10.1016/j.jcbs.2020.07.002

Stevens, J. R., Rosati, A. G., Ross, K. R., & Hauser, M. D. (2005). Will travel for food: Spatial discounting in two new world monkeys. *Current Biology*, *15*(20), 1855-1860. https://doi.org/10.1016/j.cub.2005.09.016

Stewart, I., & Lavelle, N. (2013). Auditory stimulus equivalence and non-arbitrary relations. *The Psychological Record*, *63*(3), 409-425. https://doi.org/10.11133/j.tpr.2013.63.3.001

Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2002). Stimulus equivalence and non-arbitrary relations. *The Psychological Record*, *52*(1), 77-88. https://doi.org/10.1007/BF03395415

Tyndall, I. T., Papworth, R., Roche, B. T., & Bennett, M. (2017). Differential effects of word-repetition rate on cognitive defusion of believability and discomfort of negative self-referential thoughts post-intervention and at one-month follow-up. *The Psychological Record*, *67*(3), 377-386. https://doi.org/10.1007/s40732-017-0227-2

Vervoort, E., Vervliet, B., Bennett, M., & Baeyens, F. (2014). Generalization of human fear acquisition and extinction within a novel arbitrary stimulus category. *Plos One*, *9*(5), e96569. https://doi.org/10.1371/JOURNAL.PONE.0096569

Vilardaga, R., Hayes, S. C., Levin, M. E., & Muto, T. (2009). Creating a strategy for progress: A contextual behavioral science approach. *The Behavior Analyst*, *32*(1), 105-133. https://doi.org/10.1007/BF03392178

Villatte, M. (2020). Using clinical RFT to enhance ACT interventions: The example of values work. In M. E. Levin, M. P. Twohig, and J. Krafft (eds.), I*nnovations in Acceptance and Commitment Therapy: Clinical Advancements and Applications in ACT* (pp. 30-40). Context Press.

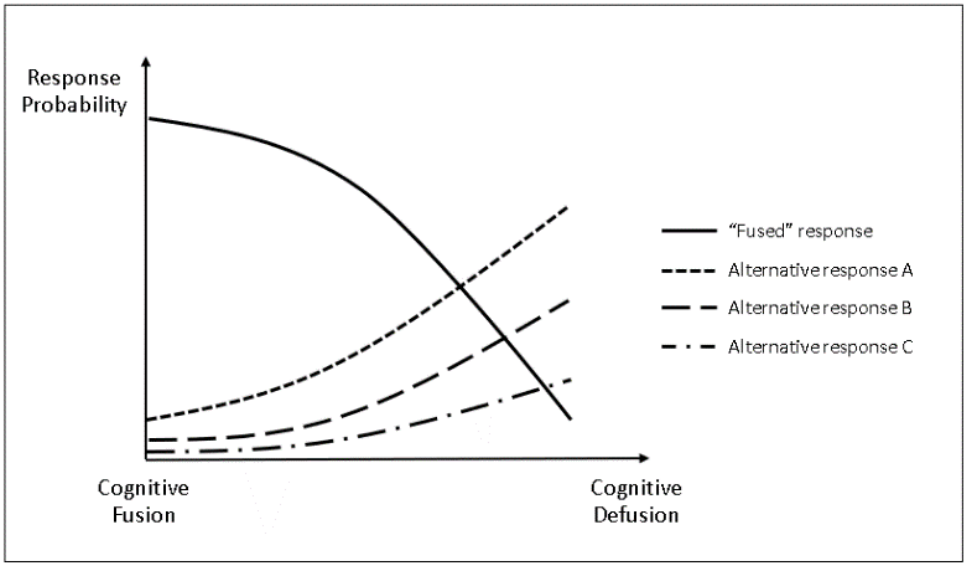
Vollmer, T. R., & Iwata, B. A. (1992). Differential reinforcement as treatment for behavior disorders: Procedural and functional variations. *Research in Developmental Disabilities*, *13*(4), 393-417. https://doi.org/10.1016/0891-4222(92)90013-V

Zhang, C. Q., Leeming, E., Smith, P., Chung, P. K., Hagger, M. S., & Hayes, S. C. (2018). Acceptance and commitment therapy for health behavior change: a contextually-driven approach. *Frontiers in Psychology*, *8*, 2350. https://doi.org/10.3389/fpsyg.2017.02350

## Figures and Tables

### Figure 1.

Visual representation of response probability in the states of cognitive fusion and defusion




### Figure 2.

Schematic representation of possible pathways to promote cognitive defusion

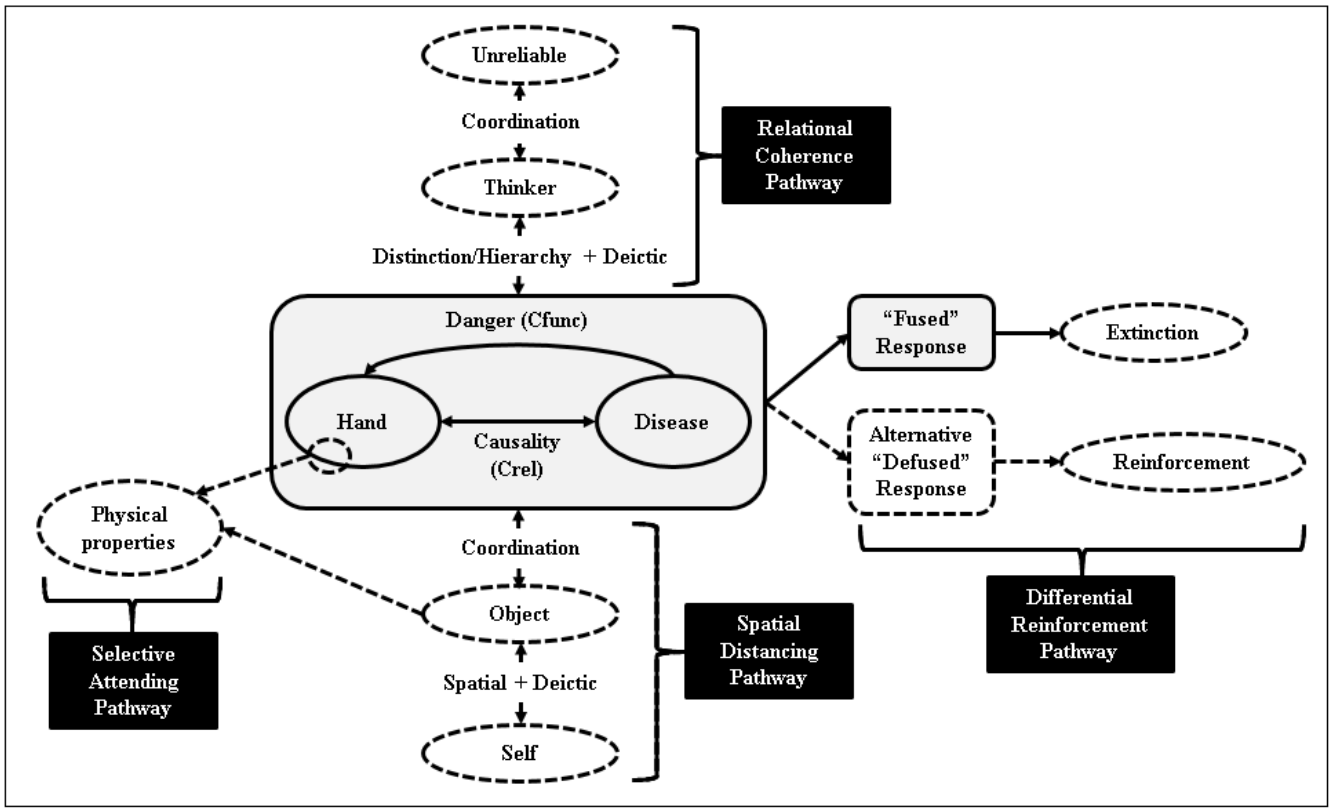


**Table 1.**

*Conceptualization of cognitive defusion*

| Initial state | Procedure | ACT exercises | Possible processes | Final state (outcome) |
|---|---|---|---|---|
| Verbally transformed functions exert strong control over behavior to the detriment of other sources of influence (i.e. cognitive fusion) and is causing psychological suffering by evoking problematic responses or failing to evoke appropriate responding | Increase salience of formal properties | Playing with words | Selective attending | Broadening of stimuli control, with reduced control of previously established verbal functions and increased control of other stimuli sources without altering the symbolic relations between stimuli (i.e. cognitive defusion) |
| | Arbitrarily attribute physical properties | Objectifying thoughts | Selective attending | |
| | Spatial and deictic framing between self and thoughts | | Spatial distancing | |
| | Establish network relating the thought with the speaker; and the speaker as unreliable | Distinguishing thought from thinker<br>Understanding thinker's function | Relational coherence | |
| | Differential reinforcement of alternative behavior (DRA) | Evaluating functional utility<br>Disobeying thoughts | Reinforcement<br>Extinction | |