The Moderating Impact of Distal Regularities on the Effect of Stimulus Pairings: A Novel Perspective on Evaluative Conditioning

Sean Hughes, Jan De Houwer and Dermot Barnes-Holmes

Ghent University

Authors Note

SH, JDH and DBH, Department of Experimental Clinical and Health Psychology, Ghent University. The preparation of this paper was supported by a Government of Ireland Research Fellowship to SH, Ghent University Methusalem Grant BOF09/01M00209 to JDH, the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33), and an Odysseus I grant of the Research Foundation – Flanders (FWO) to DBH. The authors would like to thank Pieter Van Dessel, Anne Gast and Klaus Fiedler for their comments on an earlier draft of the paper. Correspondence concerning this article should be addressed to sean.hughes@ugent.be.

Abstract

Throughout much of the past century psychologists have focused their attention on a seemingly simple question: how do people come to like or dislike stimuli in the environment? Evaluative Conditioning (EC) - a change in liking due to the pairing of stimuli - has been offered as one avenue through which novel preferences may be formed and existing ones altered. In the current article, we offer a new look at EC from the perspective of Contextual Behavioral Science (CBS) and, more specifically, Relational Frame Theory (RFT). We briefly review the EC literature, introduce Contextual Behavioral Science (CBS), Relational Frame Theory (RFT), and then describe a behavioral phenomenon known as arbitrarily applicable relational responding (AARR). Afterwards, we examine the relationship between EC and AARR. This novel perspective offers ways to organize existing as well as predict new EC effects, contributes to debates on “genuine” EC, human versus non-human EC, and further facilitates the development and refinement of cognitive theories of EC.

Keywords: Evaluative Conditioning, Relational Frame Theory, AARR

The Moderating Impact of Distal Regularities on the Effect of Stimulus Pairings: A Novel Perspective on Evaluative Conditioning

Although humans may be biologically prepared to prefer certain stimuli over others, many of our likes and dislikes are learned through on-going interactions in and with the environment (De Houwer, 2007). These evaluations are thought to play a causal role in a diverse spectrum of behaviors such as consumer choices (Gibson, 2008; Hollands, Prestwich & Marteau, 2011), voting intentions (Galdi, Arcuri & Gawronski, 2008), in-group favoritism, and stigmatization (Walther, Nagengast & Trasselli, 2005), to mention just a few. In order to understand, predict, and influence these behaviors in a sophisticated manner, researchers have sought to identify the factors responsible for the formation and change of evaluative responses.

Evaluative responses towards stimuli in the environment can be established in a wide variety of ways, from mere exposure (Bornstein, 1989), to socialization (Pettigrew & Tropp, 2006), descriptive information (Rydell & McConnell, 2006), and category membership (Pinter & Greenwald, 2004). However, many psychologists have focused on Evaluative Conditioning (EC) as a means of establishing and manipulating likes and dislikes (see Gast, Gawronski, & De Houwer, 2012). Broadly speaking, EC refers to a change in liking that is due to the pairing of stimuli. Typically, a neutral stimulus acquires the valence of a positive or negative stimulus with which it was previously paired. For example, contiguous presentations of an unknown Pokémon character with pleasant images often results in that character being rated positively whereas pairing it with negative images results in it being rated negatively (Olson & Fazio, 2001).

Over the past several decades, researchers from two intellectual traditions have studied changes in evaluative responding due to stimulus pairings from their respective scientific perspectives. On the one hand, the vast majority of this work has been conducted by social and cognitive psychologists interested in the mental processes and representations that mediate the impact of stimulus pairings on liking (i.e., *the mental level of analysis*; see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for a recent review). On the other hand, and unknown to much of psychological science, researchers who describe themselves as contextual behavioral scientists and/or behavior analysts have also studied a range of behavioral effects that seem directly relevant to EC (e.g., Smyth, Barnes-Holmes, & Forsyth, 2006; Valdivia-Salas, Dougher, & Luciano, 2013). Unlike their cognitive counterparts, however, these researchers make no appeal to, or assumptions about, mental constructs or their causal agency in behavior and its change. Rather, they aim to determine what factors in the environment influence our behavior, including those behaviors that seem to be involved in liking and disliking specific stimuli (i.e., *the functional level of analysis*; see De Houwer, 2011a)[1].

Unsurprisingly, developments in the cognitive EC literature have very rarely filtered into and informed progress in CBS, and vice-versa, because both communities draw upon radically different sets of scientific goals, values, and assumptions. Yet we believe that there is much to be gained and little to be lost by fostering greater communication between these two traditions. Cognitive researchers willing to venture into the CBS literature will find a previously undiscovered country populated with novel procedures, conceptual analyses, and findings that may contribute to their understanding of human likes and dislikes. Similarly, contextual behavioral scientists may acquire new ideas, procedures, and information through active and full engagement with their cognitive counterparts. In the current paper, we hope to showcase just what can be achieved when the functional approach to EC (adopted by CBS) interacts with the mental level of analysis and vice-versa.

[1] The relationship between contextual behavioural science (CBS) and behaviour analysis is quite complex and a general consensus on the nature of that relationship has yet to emerge. However, for ease of communication in the current article we will simply use the acronym “CBS” to refer to the broad group of researchers who conduct their scientific work at the functional level of analysis.

In Part I we begin with a brief introduction to the EC literature (for a more comprehensive review and meta-analysis see Hofmann et al., 2010). This overview will clarify how empirical and theoretical attention in EC research has largely been situated at the mental level of analysis, with changes in liking (behavior) that occur due to the pairing of stimuli (environment) explained in terms of mediating constructs such as mental associations or propositions. In Part II, we turn our attention to CBS and examine how the functional approach adopted in this tradition differs from the functional approach that is typically employed throughout the EC literature. Differentiating between these two functional approaches will require that we first consider the philosophical underpinnings and scientific goals of the former (CBS) and the various ways in which it deviates from the latter. In Part III, we introduce a functional theory of human language and cognition known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). At its core, RFT is concerned with a behavioral phenomenon known as *arbitrarily applicable relational responding* (AARR) which, broadly speaking, refers to the ability to act as if stimuli are related to one another (a) regardless of their physical properties, and (b) in the absence of direct training or instruction. Evidence indicates that, once learned, the ability to AARR changes how humans interact with, and adapt to, their physical, social, and verbal environments.

In Part IV, we draw upon the EC and RFT literatures to put forward a novel perspective on EC. First, we argue that stimulus pairings may - in principle - function as a mere *proximal cause* of changes in liking. The expression “mere proximal cause” entails that stimulus pairings only lead to a change in liking because there is a regularity in how stimuli are presented in space and time in the current (i.e., proximal) situation. Second, we argue that a more likely scenario is one in which distal regularities (i.e., regularities in the past environment of the organism) moderate the impact that proximal regularities (stimulus pairings) have on changes in liking (i.e., EC as an instance of moderated learning). Third, there may be many types of (proximal and distal) regularities that can potentially moderate the impact that stimulus pairings have on liking. Yet we believe that a particular set of distal regularities (those which give rise to the ability to AARR) cause stimulus pairings to be transformed from a mere proximal cause of liking to a *proximal cue* signaling that the CS and US are related in some way. The specific relationship between the CS and US that pairings signal will determine the properties of the changes in liking that are observed. We showcase how this new perspective may accelerate research on and theorizing about EC. For instance, our approach equips researchers with a means to systematically organize different types of EC effects so that their similarities and differences are made evident (i.e., it has heuristic value). It may also lead to the discovery of novel and unexpected ways in which liking can be changed via stimulus pairings (i.e., it has predictive value). This perspective also provides valuable input into debate on what can be considered as 'genuine' EC and raises the question of how EC in humans is related to EC in non-human animals. Finally, we argue that CBS in general, and RFT in particular, offer a language capable of describing the entire gamut of EC (and related) effects in purely functional, non-mental terms. Adopting this language may facilitate the development and refinement of cognitive theorizing in several important ways.

Part I: EC as Procedure, Effect and Mental Process

The ease and versatility with which likes and dislikes can be formed, manipulated, and eliminated via the pairing of stimuli has led to a wave of EC research that has swept through many areas of psychological science, including health psychology (Hollands et al., 2011), consumer psychology (Gibson, 2008), social psychology (Walther et al., 2005), and clinical psychology (Houben, Schoenmakers, & Wiers, 2010). This work has offered valuable insight into the *procedures* that give rise to a change in liking when stimuli are paired, the *effects* generated by those procedures as well as the mediating mental *processes* assumed to govern changes in liking resulting from the pairing of stimuli (De Houwer, 2007). In what follows we consider how recent empirical developments have shaped our understanding of EC in each of these respective areas.

Evaluative Conditioning as a Procedure

When defined as a procedure, EC refers to ways of arranging the environment so that stimuli are paired and changes in the liking of those stimuli can be observed (De Houwer, 2007, 2011b). In a prototypical EC study, a neutral stimulus (often referred to as a conditioned stimulus; CS) is repeatedly paired with a positive or negative unconditioned stimulus (US) and it is determined if these pairings lead to a change in liking. Such procedures are typically considered to be a specific subclass of Pavlovian conditioning (PC) procedures, the only difference being that EC procedures focus on changes in evaluative responses whereas PC procedures can capture any type of behavioral change (De Houwer, 2007). Furthermore, whereas the term PC is typically reserved for procedures that involve biologically significant USs (e.g., food or electric shocks), in most EC studies, neither CS nor US are biologically significant.

Broadly speaking, EC procedures may involve the pairing of neutral stimuli such as fictitious consumer products (Pleyers, Corneille, Luminet & Yzerbyt, 2007), cartoon characters (Olson & Fazio, 2001), nonsense words (Stahl & Unkelbach, 2009), and unknown individuals (Hütter, Sweldens, Stahl, Unkelbach & Klauer, 2012) with valenced words (Walther, Langer, Weil & Komischke, 2011) or images (Corneille, Yzerbyt, Pleyers, Mussweiler, 2009) with the aim of producing a change in liking. These procedures are not restricted to the use of visual stimuli but may also involve gustatory (Gast & De Houwer, 2012), olfactory (Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005), tactile (Hammerl & Grabitz, 2000), and auditory stimuli (van Reekum, van den Berg & Frijda, 1999). Although changes in liking generated by such procedures have often been indexed via self-reported ratings, indirect tasks such as evaluative priming (Fazio, Jackson, Dunton, & Williams, 1995;

Wittenbrink, Judd, & Park, 1997), the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), and the Affective Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) have allowed automatic evaluative responses to be captured and subjected to empirical scrutiny (see Nosek, Hawkins & Frazier, 2011). Although these latter procedures typically make use of reaction-time or accuracy based performances, EC can - in principle - also be studied using physiological and neurological measures of evaluation (e.g., Klucken, Kagerer, Schweckendiek, Tabbert, Vaitl & Stark, 2009; see De Houwer & Moors, 2010, for a discussion).

Evaluative Conditioning as an Effect

EC as an effect refers to the observed change in liking that is due to the pairing of stimuli. Hence, it involves more than the mere pairing of stimuli or the observation of a change in liking. This extra element is the presence of a causal relation between the two elements. More specifically, the claim that an EC effect has occurred implies that the observed change in liking was *determined by* the pairing of the stimuli. When defined as an effect, similarities and differences between EC and other evaluative learning effects become clear. It has been argued, for example, that EC differs from other types of evaluative learning with regard to the regularity in the environment that produces the change in evaluative responding (De Houwer, 2007; De Houwer et al., 2013). For instance, changes in liking that are due to the successive presentation of a single stimulus (i.e., the mere exposure effect; Bornstein, 1989) or to response-dependent contingencies (i.e., approach/avoidance learning; Kawakami et al., 2007) are typically not considered to be instances of EC.

We now know that changes in liking due to the pairing of stimuli are sensitive to the order, number, and timing of stimulus presentations (e.g., Bar-Anan, De Houwer, & Nosek, 2010; Jones, Fazio, & Olson, 2009; Stahl & Unkelbach, 2009), the manner in which the CS-US relation is established (e.g., sensory preconditioning, higher-order conditioning; Hammerl

& Grabitz, 1996; Walther, 2002), the sources of contextual control (e.g., discriminative stimuli; Baeyens, Crombez, De Houwer, & Eelen, 1996; Gawronski, Rydell, Vervliet, & De Houwer, 2010), post-acquisition modifications to the CS-US contingency (e.g., extinction, counter-conditioning, US-revaluation; Hofmann et al., 2010; Kerkhof, Vansteenwegen, Baeyens, Hermans, 2010; Walther, Gawronski, Blank, & Langer, 2009) as well as the type of organism that is tested (e.g., psychology student, child or non-human; Boakes, Albertella, & Harris, 2007; Field, 2006; Fulcher, Mathews, & Hammerl, 2008). Although the majority of this research has directly related the CS and US on the basis of spatio-temporal contiguity, changes in evaluative responding have also been obtained as a result of observation (Baeyens, Eelen, Crombez, & De Houwer, 2001), written narratives (Gregg, Seibt, & Banaji, 2006), inferences (Gast & De Houwer, 2012), and verbal instructions (De Houwer, 2006; Balas & Gawronski, 2012).

Evaluative Conditioning as a Mental Process

To date, the EC literature has almost exclusively been guided by researchers operating at the mental level of analysis. When approached from this perspective psychological events are conceptualized as being similar to a machine, composed of discrete parts that interact and are subject to specific operating conditions (e.g., Chiesa, 1992; Bechtel, 2008). The primary scientific goal at this level of analysis is to identify the mental mechanism(s) that mediate between input (environment) and output (behavior). The researcher's role is to develop an account of mental representations and processes that mediate changes in behavior. The truth or scientific value of a mental model is therefore based on the *correspondence* between the mental mechanism it proposes and the set of behavioral observations that it aims to predict. Put another way, research at the mental level of analysis is focused primarily on the prediction of behavioral effects through the use of theoretical models that bridge environmental events and behavioral outcomes.

When applied to EC, this approach assumes that mental processes and representations mediate between the pairing of stimuli and the observed change in liking. The researcher's goal is to postulate a mental theory that can explain (a) how the pairing of stimuli produces a change in evaluative responding, and (b) why the magnitude and direction of such responses are dependent on specific properties within the wider environment. Broadly speaking, the universe of mental accounts of EC can be sub-divided into two overarching sets. The first and currently dominant position is that the pairing of stimuli results in the automatic "bottom-up" formation of mental associations in memory. For example, Baeyens, Eelen, Crombez, and Van den Bergh (1992) argued that CS-US pairings result in the formation of an association between the representation of the CS and the representation of the US in memory. These changes at the representational level are thought to emerge in the absence of conscious awareness, attentional resources or the intention to relate the CS and US (i.e., they demonstrate many of the features of automaticity; see Jones et al., 2009, and Martin & Levey, 1994, for related accounts). Broadly speaking, once the association has been formed, subsequently encountering the CS will result in the automatic activation of the (valenced) US representation which in turn leads to a change in liking of the CS.

A second class of mental models have recently emerged that reject the associative position and propose that most if not all forms of classical conditioning - including EC - arise due to the formation of propositions about the CS-US relation (Mitchell, De Houwer, & Lovibond, 2009). Whereas mental associations are simply hypothetical structures in memory through which activation can spread between representations, propositions are qualified truth statements about the environment. As statements about the environment, they can be true or false (i.e., correspond to the actual environment or not) and can specify not only that stimuli in the environment are related but also how those elements are related (e.g., "*the CS is opposite to the US*"). According to this propositional account, participants utilize their knowledge of how stimuli are related as the basis upon which to evaluate the CS. These propositions can be acquired in a number of ways, from prior knowledge and direct experience, to verbal instructions and deductive reasoning (De Houwer, 2009). The formation of these propositions is assumed to require an awareness of the stimulus relation, as well as the time, cognitive resources and intention to relate those stimuli[2].

Part II: Contextual Behavioral Science

In this section, we introduce CBS, and consider how research conducted within this tradition, particularly from the perspective of RFT, may provide an alternative approach to the study of EC. Before proceeding, it is worth considering several points. First, we recognize that for cognitive and social psychologists, the non-mental approach discussed below represents a significant departure from their own pre-analytic assumptions and scientific aims. Indeed, many readers may find the philosophy of science (functional contextualism) and theoretical perspectives (RFT) from which contextual behavioral scientists operate to be entirely new territory. It therefore seems prudent to provide some background information about both of these topics so that the reader can better appreciate the nature of CBS, how it differs from the cognitive approach - and perhaps most importantly - how it may contribute to EC research and cognitive theorizing. At the same time, researchers might argue that they have already adopted a functional approach when it comes to the study of EC. Although there may be some substance to this argument, we will attempt to show that the functional (*effect-centric*) approach adopted within psychological science is very different to the functional (*analytic-abstractive*) approach found within CBS. Furthermore, we contend that the latter functional approach offers EC researchers possible advantages that the former functional approach does not.

[2] From a mental (mechanistic) perspective EC effects can - in principle - be mediated by any type or combination of mental processes. Thus while many researchers subscribe to either an associative or propositional position, others argue that EC effects can be driven by both processes operating singularly, or in interaction, in an automatic or non-automatic fashion (e.g., Gawronski & Bodenhausen, 2011).

Contextual Behavioral Science

While various forms of functionally oriented psychologies have emerged over the last century the most empirically and theoretically productive (contemporary) branch is arguably that of CBS. At the core of this tradition resides a philosophy of science known as functional contextualism which specifies (a) the assumptions, goals and values of the researcher, and by implication, (b) the principles, theories and methodologies that they draw upon (for more on CBS and how it relates to radical behaviorism see Hayes, Barnes-Holmes, & Wilson, 2012). As a contextualist behavioral scientist sees it, (psychological) science involves a single unified goal: to predict-and-influence behavior with precision (applying a restricted set of principles to any event), scope (explain a comprehensive range of behaviors across a variety of situations), and depth (cohere across analytical levels and domains such as biology, psychology, and anthropology). In what follows we highlight how a CBS approach differs from the mental (mechanistic) approach to psychological science.

Absence of mental mediators. CBS adopts an exclusively functional epistemology that makes no appeal to mental mediators. Instead the functional relationships between environment and behavior (which unfold across time and context) are treated as causal or explanatory factors. To illustrate this point more clearly, consider the phenomenon of latent learning in which experiences at an earlier point in time do not influence responding at that point but instead only impact behavior at a later date (e.g., Diaz & Gonzalo De la Casa, 2002; Lubow & Gewirtz, 1995). If one operates at the mental level of analysis then this change in behavior is viewed as the product of some intervening (mental) mechanism that mediates between the experienced event at time 1 and the observed response at time 2. However, from a CBS perspective, it is enough to say that behavior at time 2 is a function of experience at time 1 (regardless of the delay between an event and behavior) provided that this functional relation allows the researcher to predict-and-influence that behavior with precision, scope and depth (see also De Houwer, Barnes-Holmes, & Moors, 2013).

In other words, appeals to - or assumptions about - mental constructs or their causal agency in producing behavior is a strategy that contextual behavioral scientists choose not to adopt. Rather, they define behavior as an on-going act that always occurs within, and in response to, a current and historical context. This ‘act-in-context’ can vary from the most proximal behavioral instance (e.g., evaluative responses to stimuli in the current environment) to temporally distal and remote behavioral sequences (e.g., the impact of a particular experience two years ago on choosing whether to sit beside a member of another racial group at the cinema). Given that the temporal and spatial parameters of a context can vary dramatically it can be difficult to determine which contextual framing of an act is “true” in the sense of “most correct”. CBS researchers therefore adopt a ‘pragmatic truth criterion’ that qualifies the success, meaning or validity of a scientific analysis in terms of its ability to achieve prediction-and-influence, with precision, scope, and depth over the behavior of interest. Thus, in CBS, the study of EC effects involves a shift away from the search for mental mediators of EC (e.g., associations or propositions via which pairings produce changes in evaluative responding) and towards the identification of environmental moderators of EC (e.g., stimuli or events that determine whether pairings have a particular impact on evaluative responding).

The challenge of predicting-and-influencing behavior. The choice of functional researchers not to focus on mental mediators follows, in part, from their aim to predict-and-influence behavior. Consider, for example, the mental association models of EC outlined above, wherein a change in liking due to the pairing of stimuli is mediated by the formation, activation, or modification of mental associations in memory. As pointed out by Gardner (1987), these mental entities may be treated as the cause of a particular behavior (e.g., evaluative associations as mental causes of ratings on a Likert scale) but they are not open to direct manipulation. For instance, there appears to be no method with which to interact directly with mental constructs such as associations or propositions. Instead (experimental) researchers must (a) act on the world in some way, such as presenting stimuli to an organism, and (b) observe a change in behavior in that organism. Based on the effect of their actions on behavior, researchers can then postulate a mediating mechanism that is responsible for the obtained change in behavior or outcome.

This ‘manipulation argument’ is not intended to dismiss or undermine analysis at the mental level, nor does it imply that mental constructs cannot be studied in a scientific manner. Rather, the take home message is that one’s preference for using (or not using) mental constructs in scientific analyses depends on the aim of one’s scientific endeavor. If a researcher’s analytic goal involves predicting behavioral *effects* under certain environmental conditions, then associations, propositions, emotions, or any other mental or non-mental variable (e.g., neurological activity) can be used to achieve prediction, provided that it reliably precedes that effect. On the other hand, if a researcher’s primary objective is to *predict-and-influence* behavior, then focusing exclusively on mental explanations is ultimately insufficient and could be deemed as distracting the researcher from the goal at hand. In order to exert influence over behavior the researcher must successfully manipulate events external to that behavior, and only contextual variables can be directly manipulated; mental variables cannot (see De Houwer, 2011a; Hayes & Brownstein, 1986). Consequently, from a CBS perspective, scientific analysis is not complete until the causal variables external to the behavior of interest have been identified – not because of some dogmatic adherence to a physical monism that excludes the non-physical, mental world, but rather as a means to achieve its scientific goals (for more on CBS see Hayes et al., 2012).

Development of behavioral principles and functional theories. The development and evaluation of theories within CBS also differs from that at the mental level of analysis. While accounts at the mental level involve the postulation of mental processes or representations in order to understand past and predict future behavior, functional theories are different: they start out by identifying (a) specific functional relationships between behavior and the environment, and then (b) abstract these relationships into overarching ‘behavioral principles’ that are high in precision, scope and depth. Examples include the principles of reinforcement, punishment, stimulus generalization, and discrimination. These principles are primarily inductive in nature, built from the bottom up, and “apply across a broad array of topographically distinct behaviors of varying complexity while maintaining coherence and parsimony” (Levin & Hayes, 2009, p.6). Functionally oriented researchers treat these principles as their primary analytic tools and draw upon them to study and evaluate behavior. For example, the development of temper-tantrums in children, emergence and change of problematic behaviors in pets, as well as the tendency to check one’s phone upon hearing a ringtone, can be accounted for by drawing upon the notion of reinforcement (see Catania, 1997). Likewise, stopping one’s car at a red traffic light and accelerating in the presence of a green light or taking a cake out of an oven after an alarm rings can be explained in terms of stimulus control. Analyzing an individual’s interactions with the environment in terms of these behavioral principles is often defined as a carrying out a ‘functional analysis’.

Theories at the functional level emerge when researchers seek to move from functional analyses of individual behaviors to more general phenomena such as language, reasoning and cognition. Such theories constitute organized sets of interrelated behavioral principles. These models or theories are considered to be ‘true’, successful, meaningful or valid to the extent that as they ‘work’ (i.e., they allow the researcher to achieve prediction-and-influence over the phenomenon of interest). Thus the functional level of analysis is richly theoretical. But it is a type of theorizing that differs markedly from that seen elsewhere in psychological science (see Barnes-Holmes & Hussey, 2015; Hayes, 1996, 1998; Wilson, Whiteman, & Bordieri, 2013).

Functional Level of Analysis: Analytic-Abstractive vs. Effect-Centric.

One could argue that many psychologists adopt a functional approach both inside and outside of CBS. For instance, all experiments involve the manipulation of independent variables whose effect on dependent variables is examined. What should now be clear, however, is that the functional approach adopted in CBS differs markedly from the functional approach seen in many other areas of psychological science. The former begins by conducting functional analyses of specific environment-behavior interactions (e.g., temper tantrums, disobedient pets) and then abstracts these relations into overarching behavioral principles (e.g., reinforcement) which apply to a wide variety of situations and behaviors (i.e., it is *analytic-abstractive*). The functional perspective adopted in psychological science is typically of a different kind: the description of individual environment-behavior relations with the aim of predicting the behavioral effect obtained from a certain procedure or set of procedures (i.e., it is *effect-centric*).

These two functional approaches, and the analyses that they engender, differ in a number of important ways. Whereas an analytic-abstractive approach strives to explain a wide variety of seemingly unrelated behaviors using a restricted number of behavioral principles (i.e., it aims for high *scope* and *precision*) an effect-centric approach is typically interested in only those aspects of the current environment or procedure that give rise to a specific or narrow range of behavioral effects (i.e., it emphasizes *precision* over *scope*). Consequently, while the former approach provides a means to conceptualize seemingly unrelated behaviors as instances of the same underlying phenomenon (i.e., it has heuristic value), the latter does so to a much lesser extent. To illustrate, take the previous examples of temper tantrums in children, a disobedient pet and persistent checking of one's phone. Approaching these behaviors from an effect-centric position might lead researchers to postulate three separate effects (e.g., a temper tantrum effect, a disobedience effect, and a checking effect) and suggest that, because those behaviors look different, involve different organisms, stimuli and events, they must reflect three separate and unrelated phenomena. Yet from an analytic-abstractive position these three types of behavior can be viewed as different instances of the same underlying phenomenon (operant learning). Similarly, an effect-centric perspective might lead researchers interested in cognitive control to view outcomes such as the Stroop effect, Simon effect, and Task-Rule Congruency effect topographically, rather than as different instances of the same underlying behavioral principle (stimulus control; Liefooghe & De Houwer, in press). Although these different effects are sometimes related at non-functional levels (e.g., at the mental level by assuming that the same mental processes mediate different effects; e.g., Miyake et al., 2000), they remain distinct within an effect-centric functional approach. Returning to the topic of evaluative conditioning, an effect-centric approach might view changes in liking due to the pairing of stimuli via instructions, observation, inferences or spatio-temporal contiguity as entirely different phenomena given that they are instantiated in ways that, at least on the surface, appear to differ from one another. Once again an analytic-abstractive perspective might suggest that these topographically distinct outcomes may stem from a common functional source (we will return to this issue later in the paper).

Effect-centric and analytic-abstractive approaches also differ in terms of the emphasis that they place on *distal* regularities and their impact on proximal regularities. Effect-centric approaches often appear to ignore or play down the learning histories that likely gave rise to specific capacities associated closely with the human species, such as complex language skills and advanced reasoning (for related arguments see Fiedler, 2014). Instead, an effect-centric approach typically focuses on those *proximal* regularities which need to be manipulated in order to produce the desired outcome (e.g., pairing stimuli on a computer screen). Although certain moderators of this effect - such as the context, stimuli, or organism - can also be manipulated, more distal regularities and learning histories are often taken for granted or entirely ignored. This may result in a disconnect between past and present events. For instance, imagine that the word ‘vomit’ is repeatedly paired with a neutral face and a change in liking is observed. Or that the word ‘loathes’ is presented along with pictures of neutral faces and aggressive looking dogs. Researchers may not be interested in the learning history that allows the word ‘vomit’ to function as a US or how the word loathes acquires its ability to transform the relation between neutral and valenced stimuli. The fact that such stimuli give rise to and influence changes in liking is enough for the researcher to achieve their scientific goal of predicting EC effects. Yet, as outlined above, such a stance is unacceptable for researchers interested in achieving both prediction-and-influence over behavior. It is only by identifying the learning processes that underpin the very abilities to engage in complex language and reasoning that we may (ultimately) influence the above changes in liking in a sophisticated manner. Therefore an analytic-abstractive approach emphasizes the importance of an individual’s learning history (or distal regularities) and the impact they have on proximal regularities.

The take home message here is that adopting an effect-centric versus analytic abstractive functional approach will have important consequences for the researcher, and by implication, the perceived value they place on a functional level of analysis. Those who adopt an effect-centric position will tend to generate ever growing piles of behavioral effects that are typically described in purely topographical terms or linked narrowly to specific procedures in the current context (see Meiser, 2011). This in turn may lead to the view that many effects are unrelated to or somehow independent from others, and as a result commonalities between different outcomes are unlikely to be identified. The end result of a purely effect-centric functional approach is a deeply unsatisfying one: an accumulating set of behavioral outcomes that lack a conceptual framework capable of systematizing such outcomes or making novel predictions. In contrast, the analytic-abstractive approach adopted within CBS offers a more sophisticated functional perspective: One capable of identifying functional relations between the environment and behavior and then *abstracting* those relations into overarching behavioral principles and theories that apply to both specific (precision) and general (scope) behaviors. In other words, the functional analytic-abstractive approach is focused on organizing and systematizing seemingly unrelated findings in the immediate service of increasing prediction-and-influence over the phenomena of interest. And it attempts to do this by drawing attention to the dynamic interplay between distal and proximal regularities in the environment.

Strangely, this distinction between effect-centric and analytic-abstractive approaches has rarely, if ever, been made in psychological science or within the EC literature (although see Barnes-Holmes & Hussey, 2015). Instead, the prevailing norm is to treat the functional level as synonymous with the former rather than latter perspective (e.g., Proctor & Urcuioli, in press). It therefore comes as no surprise that many cognitive and social psychologists consider the functional (effect-centric) level as unsustainable by itself because it leads to a proliferation of behavioral effects in the absence of any theoretical framework capable of organizing those findings or making novel predictions (Meiser, 2011). We agree. It also comes as no surprise that they choose to devise frameworks to systematize findings and do so at the mental level of analysis given their own scientific goals, values, and assumptions. This is not a problem. Yet to equate the functional level with an effect-centric approach *is* a mistake. The analytic-abstractive position outlined here and adopted within CBS circumvents many of the perceived issues with an effect-centric approach and offers new insight into much of human psychological life (for recent reviews see Hughes & Barnes-Holmes, in press-a,b).

In what follows we showcase how an analytic-abstractive theory known as Relational Frame Theory has done just that. Over the last twenty-five years this account of human language and cognition has sparked a flurry of conceptual and empirical work and has come to dominate a considerable portion of the CBS literature. In Part III we outline the core features of this analytic-abstractive theory and then in Part IV examine how it can be used to accelerate empirical and theoretical developments in the domain of EC.

Part III: Relational Frame Theory

At its core, RFT argues that organisms throughout the animal kingdom can learn in many different ways, from the pairing of stimuli (classical conditioning) to the relationship that exists between particular responses and its consequences (operant learning). However, early on in their lives humans develop an “advanced” type of operant behavior known as arbitrarily applicable relational responding (AARR). RFT argues that this type of behavior is something special, the basic “building block” from which human language and cognition spring forth. Put simply, ‘relational responding’ refers to a type of learned operant behavior that involves *responding to one event in terms of another*. Although humans and non-humans can both respond to the relationship between stimuli and events, the latter quickly develop a more complex type of behavior (AARR) that fundamentally alters how they interact with the world around them. In the following section we examine how RFT carves this type of operant behavior into two different varieties (non-arbitrarily and arbitrarily applicable).

Non-Arbitrarily Applicable Relational Responding (NAARR)

Mammals, birds, fish and insects can all be trained to respond to the relationship between stimuli in the environment. However, for many different species these relational responses appear to be characterized by two key properties: (1) they are rooted in a prior history of direct experience, and (2) they are defined by the physical features of the to-be-related stimuli (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001; Harmon, Strong & Pasnak, 1982; Reese, 1968; Vaughan, 1988). RFT refers to this type of behavior as an instance of non-arbitrarily applicable relational responding (or NAARR) because the organism is relating stimuli based on their formal or physical properties. Properties such as color, shape, quantity, and size are considered ‘non-arbitrary’ because they are based on the physical characteristics of the stimuli, unlike arbitrary or arbitrarily applicable properties, which are largely determined by social convention (e.g., the meaning of letter-strings such as ‘honesty’, ‘beauty’ or ‘freedom’).

To illustrate the concept of NAARR more clearly, imagine that a pigeon is exposed to a simple learning task in which a sample stimulus (e.g., a red circle) is presented at the top of a computer screen and two comparison stimuli (e.g., a red and a green circle) are presented at the bottom of the screen. On trials where a red circle serves as a sample stimulus, selecting the red circle from the available comparisons is reinforced. Whenever a green circle is the sample, selecting the green circle from the available comparisons is reinforced. Training continues in this way across a whole host of different colors and shapes. Once the bird is consistently correct across a large number of trials it is then presented with entirely novel stimuli (the matching of which were never directly reinforced in the past). Research suggests that the pigeon will continue to select a shape or color from the bottom of the screen that is physically identical to a shape or color at the top of the screen even when that particular response was never previously reinforced (Frank & Wasserman, 2005). Now consider a series of studies wherein adult rhesus monkeys (Harmon et al., 1982) or marmosets (Yamazaki, Saiki, Inada, Iriki, Watanabe, 2014) were trained to select the taller of two items which differed only in terms of their respective height. When subsequently presented with a previously ‘correct’ item (i.e., a stimulus that was taller than its comparison stimulus) as well as a novel item that was even taller, monkeys consistently selected the latter item despite the fact that selecting the former was reinforced at an earlier point in time. These studies, in addition to many others, suggest that animals can respond to the non-arbitrary (i.e., physical) relationship that exists between stimuli. In the above examples, pigeons related colored shapes based on their physical similarity to one another while rhesus monkeys responded to the comparative relationship between items that differed in their respective height.

Arbitrarily Applicable Relational Responding (AARR)

RFT argues that while humans and non-humans can both demonstrate NAARR, the former also display all the hallmarks of a more advanced type of relational behavior known as AARR. This behavior is assumed to arise from a history of generalized operant learning early on in our development and is characterized by three core properties known as mutual entailment, combinatorial entailment, and the transformation of function (see Hughes & Barnes-Holmes, in press-a).

Imagine, for example, that a child learns that a picture of a Dog (A) is the same as the word DOG (B), and that the word DOG is related to the sound D-O-G (C). The first property of AARR (*mutual entailment*) refers to the fact that when the child is taught that a picture of a dog (A) is the same as the word DOG (B) she will subsequently act as if the word (B) is the same as the picture (A) without any explicit training or instruction to do so. The second property of AARR (*combinatorial entailment*) refers to the additional stimulus relations that tend to emerge between two or more mutually entailed stimuli. Imagine that a child is shown three coins from a foreign currency of which she has no prior experience, and is told that "*Coin A is worth less than Coin B*" and "*Coin B is worth less than Coin C*". It is now likely that she will act in a number of novel and untrained ways that do not depend on the physical properties of those stimuli (e.g., she will act as if Coin A is worth less than C; that Coin C is worth more than A; that Coin B is worth more than A and that Coin C is worth more than B).

The third and final property of AARR (*transformation of function*) refers to the finding that once stimuli have been related to one another, the psychological properties of those stimuli may be changed in accordance with how they were related. For instance, imagine that you are told that four previously unknown brand products (A, B, C, D) are all basically the same, and that A is then given a positive valence through repeatedly pairing it with positively valenced images. Not only will A acquire a positive valence but B, C, and D will as well despite the fact that the latter were never directly paired with positive images and were all physically dissimilar to one another. Critically, this transformation of function depends on the relationship established between or among stimuli: if two neutral images (A and B) are related as opposite to one another, and A is then paired with shock, the fear arousing properties of A will not necessarily transfer to B. Rather the emotional properties of B may come to be transformed in-line with the nature of the stimulus relation; that is, B may elicit relaxation instead of a fear response (see Whelan & Barnes-Holmes, 2004).

AARR is contextually controlled. If humans have the capacity to relate stimuli arbitrarily in increasingly complex ways, one might ask why does this ability not lead to complete and utter chaos. For example, why do people not try to eat the word “apple”, lick the words “ice cream” off a page, or even swat the word “fly” from a book? RFT proposes that the way in which people relate stimuli (and transform functions through those relations) is under the control of stimuli in the environment known as contextual cues. While certain types of contextual cues specify how stimuli are related (e.g., “*A same as B*” or “*A prevents B*”), others specify the psychological properties that are transformed through those relations (e.g., “*A tastes disgusting*” or “*B feels soft*”). RFT researchers usually refer to the former as “relational cues” given that they specify how stimuli and events should be related. These cues can be used to relate stimuli in a near infinite number of ways, from relations based on equivalence or similarity (e.g., ‘*Hond is the same as dog*’; Cahill et al., 2007) to those based on opposition (e.g., '*Good is the opposite of Evil*'; Dymond, Roche, Forsyth, Whelan & Rhoden, 2008), hierarchy ('*Cat is a type of mammal*'; Gil, Luciano, Ruiz & Valdivia-Salas, 2012), comparison ('*Fruit is better than candy*'; Vitale, Barnes-Holmes, Barnes-Holmes & Campbell, 2008), deictics ('*I am not you*'; McHugh & Stewart, 2012), temporality ('*March comes before May*'; O'Hora et al., 2008) and causality ('*If X then Y*'). Importantly, these relational cues need not always be words; other properties of the environment such as sounds, symbols, shapes, background color (e.g., Gawronski et al., 2010) or verbal rules (e.g., Zanon, De Houwer, & Gast, 2012) may also function in a similar capacity.

At the same time, responding can also be controlled by "functional cues" in the environment that specify the type of psychological properties that are transformed in accordance with stimulus relations. For example, the verbal stimulus "ice cream" could in principle evoke many of the psychological properties of actual ice cream (such as its taste, smell, appearance, or its coolness) based on the equivalence relation between the word and the food-item. If, however, someone asks you to picture ice cream, the visual properties of ice cream would likely predominate. Likewise, in the sentence "imagine what ice cream tastes like," the expression "tastes like" may serve as a functional cue that is responsible for the fact that only the gustatory (and not other) functions of ice cream predominate. We now know that functional cues in the environment can influence (a) the ability of a stimulus to signal whether an outcome will occur after a given response (Dougher, Hamilton, Fink & Harrington, 2007), (b) whether a stimulus will elicit an emotional response (Barnes-Holmes, Barnes-Holmes, Smeets & Luciano, 2004), (c) whether a stimulus should be approached (Gannon, Roche, Kanter, Forsyth & Linehan, 2011) or avoided (Roche, Kanter, Brown, Dymond & Fogarty, 2008) as well as (d) whether a stimulus will elicit a sexual response (Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000). Once again, many different stimuli in the environment – above and beyond words - may serve as functional cues.

AARR is learned early on in our development. RFT argues that AARR is a type of operant behavior that is learned early on in our development through interactions with the socio-verbal community. To illustrate, imagine that you are attempting to teach your infant son how to name a number of objects around the house. You will likely begin by pointing at an item (e.g., a toy bear), uttering its name in the presence of your son and then encouraging any orientating response that he makes towards the item (i.e., hear the word *bear* → look at the bear). At the same time, you will also present the item to your son and then display or encourage appropriate responses (see the bear → say the word *bear*). Both of these interactions will take place in the presence of contextual cues - and in natural language interactions these cues typically take the form of questions such as 'What is this?' or 'What is the name of that?' In the language of RFT, you are directly reinforcing bidirectional responding to an object and its name in the presence of a contextual cue (e.g., look at the bear when hearing "bear" as well as say "bear" when you see the bear). Importantly, this bidirectional training will not stop here: you and others in the social community (teachers, friends, family) will likely engage in the same exercise with your son across a vast spectrum of different objects, from toys ('*where is your bike*'), to people ('*who is that*'), food ('*this is an apple*') properties of the environment ('*that is called the sun…what is that called?*') and do so in a wide variety of different contexts: at the park, home, at the shopping mall, school and so on. Although the particular stimuli, people and contexts change across time, the functional relation between those stimuli is always held constant: the child's relational responding is reinforced in both directions and in the presence of arbitrary contextual cues.

Hence, through multiple exemplar training, your son learns a type of generalized bidirectional responding which no longer depends on the physical features of the stimuli involved and that leads to mutually entailed responses being emitted in the absence of direct reinforcement. Now when you present him with a novel object and a word he has never encountered before (e.g., a laptop and its name) he will respond in a bidirectional manner without any reinforcement for doing so (e.g., he will point to the laptop when asked ‘where is the laptop’ and answer with ‘laptop’ when asked ‘what is this’). According to RFT, this bi-directional relation between an object and word represents an instance of mutual entailment in which stimuli are related on the basis of their arbitrary “similarity” to one another (i.e., your son has been taught to treat a word and its referent as functionally similar in certain contexts). After more training in which a series of stimuli are related (e.g., a teddy bear being related to the sound “teddy bear” which is afterwards related to the written words “teddy bear”), other patterns of behavior such as combinatorial entailment emerge (e.g., look at the teddy bear when seeing the written words “teddy bear”)[3].

Why is AARR so important? Over the past 40 years AARR has captured the imagination of functional researchers due to its symbolic, flexible, and generative properties. From the beginning they realized that this type of behavior was inherently *generative*. Providing humans with a small set of direct experiences consistently causes them to act as if those stimuli are related to one another in a staggering number of novel and untrained ways. For instance, after learning that the written word ‘poison’ (A) is the same as a picture containing an unknown symbol (B), and that the latter is the same as the spoken word ‘G-I-F’ (C), people will likely avoid consuming any items that contain images of B or that are labeled “GIF”. In other words, they will spontaneously act as if the above stimuli are related in four new ways based on two directly trained relations (i.e., that C is the same as A; A the same as C; B the same as A and C the same as B). When subsequently taught that three new stimuli

[3] For communicative purposes we have presented a simplified version of the origins, properties and implications of AARR for psychological science. For a far more nuanced perspective we encourage readers to consider Hayes et al. (2001); Törneke (2010); or Hughes and Barnes-Holmes (in press-a, b).

are to be treated as equivalent to one another (*X-Same-Y; Y-Same-Z*) people will act as if these stimuli are also related in four untrained ways, and when the first equivalence relation is related to the second equivalence relation sixteen additional relations will arise. Indeed, this exponential increase in the number untrained relations continues to grow as more and more stimuli are related, so that by the time eight stimulus relations are trained people can - in principle - act as if those stimuli are related in several thousand untrained ways. Thus AARR represents a type of behavior that rapidly accelerates learning as more and more stimuli are related.

A second reason why AARR has attracted so much attention is that it equips humans with an unparalleled degree of *flexibility* when interacting with the world around them. Once a person has learned how to respond in an arbitrarily applicable fashion they can relate any stimulus to any other stimulus in a near infinite number of ways. They can relate stimuli with no physical resemblance (like spoken words, written words, and pictures) and these relations can come to control how they subsequently respond. People can also act as if stimuli have acquired, changed, or lost their psychological properties without the need to directly contact contingencies in the environment. To illustrate, suppose that a person learns that a novel item (A) is less than a second item (B) and that B is less than a third item (C). Thereafter, B is repeatedly paired with electrical shocks. Evidence indicates that people will display greater fear towards C than B and more fear to B than A, despite the fact that C and A were never paired with shock and that none of the stimuli share any physical similarity (Dougher, Hamilton, Fink, & Harrington, 2007). Functionally speaking, it seems unlikely that this behavior is a simple case of (non-arbitrary) relational learning given that the three items did not differ along any relevant physical dimension, such as size. At the same time, it does not appear to be an instance of second or higher-order Pavlovian conditioning because the fear functions of the A and C stimuli differed substantively from the B stimulus. Thus it appears that when organisms learn to *relate* events in arbitrarily applicable ways, the possibilities of manipulating and changing the world are dramatically increased.

Summary. In short, AARR unshackles humans from the need to learn via direct experiences and rapidly accelerates their ability to flexibly adapt to the world around them. Over the past several decades functional researchers have explored both conceptually and empirically the histories of learning which gives rise to AARR, and how they appear to be critical to the development of complex human behaviors, including the ability to engage in complex language and reasoning (Hayes et al., 2001), to develop a rich and elaborate sense of self as well as advanced perspective taking capacities (McHugh & Stewart, 2012). AARR is also argued to play a key role in implicit and explicit cognition (Hughes et al., 2012), developmental disorders (Rehfeldt & Barnes-Holmes, 2009), psychopathology (Hayes, Levin, Plumb-Vilardaga, Villatte, & Pistorello, 2013), intelligence, and many other psychological phenomena (for a recent review Hughes & Barnes-Holmes, in press-b). In what follows we turn our attention to the domain of evaluative learning and consider how the concept of AARR may help reshape the basic understanding of EC[4].

Part IV: EC as an Interaction between Proximal and Distal Regularities

In this final section of the paper we introduce a novel perspective on EC that draws upon developments within both the EC and RFT literatures. At the core of this account reside two ideas. The first is that stimulus pairings may constitute a mere *proximal cause* for changes in liking. In this case, the observed changes in liking are due only to the proximal regularities in how stimuli are presented here and now in space and time. That is, the impact

[4] We realize that it may be tempting, and entirely possible to, explain the behavioral observations that are typically associated with AARR in terms of mental events. But it is important to realize that the very concept of AARR constitutes a functional-analytic abstractive explanation in terms of overarching patterns of interactions between the behavior of human beings and their physical, social and verbal environments. When we use terms such as mutual and combinatorial entailment we are referring to the specific ways in which people behave based on a particular history of learning (i.e., the way people act as if stimuli are related in novel and untrained ways). Likewise, terms such as the transformation of function refer to patterns of behavior rather than mental events: once stimuli have been related, people will act as if the psychological properties of those stimuli have changed in some predictable way.

of the stimulus pairings on liking is not moderated by distal experiences. A second possibility is that the impact of proximal regularities (stimulus pairings) on liking may be moderated by other regularities. This moderated learning perspective argues that EC effects depend on more than the mere experience of stimulus pairings and can be moderated by additional factors in the environment.

In what follows we briefly discuss these two possibilities. We then introduce a third idea that builds on the second idea, namely that a specific set of distal regularities (i.e., those that give rise to the ability to AARR) are central to many of the effects observed in the EC literature. These regularities transform spatio-temporal contiguity from a mere proximal cause to a *proximal cue* signaling that the CS and US are related to one another. The specific relationship between the CS and US that pairings signal determines the properties of the changes in liking that are observed. This position corresponds to the analytic-abstractive function approach typically adopted in the CBS literature insofar as different EC effects are viewed as instances of the same behavioral phenomenon (AARR). Once we discuss the above possibilities we then show how an "EC as AARR" account leads to new perspectives on existing EC effects, new perspectives on what constitutes "genuine" EC, clarifies differences between human and non-human EC, and also provides a novel impetus for the development of cognitive theories of EC.

EC as the Product of Proximal Regularities

A first possibility, one that is often entertained in the EC literature, is that stimulus pairings function as a mere proximal cause of liking. The expression "mere proximal cause" entails that stimulus pairings exert their influence on liking solely due to a regularity in how stimuli are presented in space and time. Although the organism may have contacted distal regularities during their past interactions with the environment, these regularities are not assumed to influence the way in which the pairings influence liking. Let us be clear here.

Stating that pairings function as a mere proximal cause of liking certainly allows for a distal history of learning experiences to alter the responses to each of the individual *stimuli* that are paired. For instance, a distal history of language learning may transform arbitrary lines on a page (e.g., the word "cancer") into a negatively valenced stimulus that might change the liking of stimuli it is paired with. But from a mere proximal cause perspective, that same history of learning is not assumed to influence the effect of the *pairings* (i.e., pairings are not thought to be an interpreted event or one that is given semantic meaning by the organism). Thus, to adopt a "pairings as a mere proximal cause" position, one has to assume that when people have acquired verbal skills, they either do not apply those skills when confronted with stimulus pairings, or apply them selectively to the individual stimuli that are paired but not to the fact that stimuli are paired. In what follows we argue that such a position is certainly possible yet highly unlikely, and that distal regularities may not only transform the meaning of individual stimuli but also the function or "meaning" of pairings as a proximal event.

EC as an Interaction between Proximal and Distal Regularities

A second possibility, and one that we alluded to above, is that the impact of proximal stimulus pairings on liking may be moderated by the organism's previous learning experiences. In-line with De Houwer et al. (2013) we define moderated learning as the effect of regularities in the environment on how other regularities in the environment influence behavior. When applied to EC, a moderated learning perspective argues that distal regularities may influence the impact that proximal regularities (stimulus pairings) have on liking. They may do so in two ways, by either influencing: (a) the way in which organisms respond to the individual stimuli that are paired and/or (b) the way in which organisms respond to pairings as a proximal event. Note that at this stage, a "moderated learning" perspective of EC does not commit to any specific theory about the development, nature, representation, or deployment of those distal learning experiences that give rise to verbal skills or language. Rather it merely appeals to the idea that humans have encountered distal regularities in the past, and as a result of these interactions with the environment, they possess and can utilize verbal skills in a flexible manner. For instance, our arguments do not depend on whether language necessarily involves propositional representations (mental level) or is an instance of arbitrarily applicable relational responding (AARRing; functional level). When we say that verbal skills influence how people respond to stimuli or pairings as a proximal event, we simply mean that EC effects involving those stimuli and pairings would either not occur or would not have the same properties for organisms that have not acquired the ability to use language (De Houwer & Hughes, submitted). Thus a moderated learning perspective on EC is - at least in the first instance - situated at a very abstract level of analysis (see De Houwer & Moors, 2015). Let us now consider just one possible way in which this moderated learning perspective could be further specified.

EC as an Instance of AARR

Earlier in this paper we described empirical (AARR) and theoretical (RFT) developments that are currently shaping the functional (*analytic-abstractive*) approach to human language and cognition. Interestingly, when these developments are combined with the aforementioned moderated learning account of EC an entirely novel perspective emerges – one in which distal regularities that give rise to AARR moderate the impact of stimulus pairings on liking. This perspective on EC is based on the idea that a particular set of regularities moderate the impact of stimulus pairings on liking (i.e., those that give rise to the ability to AARR). According to this perspective when participants enter the experimental context they bring with them a long history of relational learning that influences how they respond to stimuli that are paired in space and time. This arbitrarily applicable relational responding is not 'switched off', discarded or somehow 'set aside' when one encounters such pairings but directly informs and dictates how individuals respond to those regularities. Put simply, once people have learned how to AARR, pairings as a proximal event may be transformed from a mere *proximal cause* to a *proximal (contextual) cue* signaling that stimuli are related in some way. The relational "meaning" or function of stimulus pairings will determine the manner in which the CS and US are related, and by implication, the nature of the change in liking observed. This account of EC is analytic-abstractive insofar as different EC effects are viewed as instances of the same underlying behavioral phenomenon (AARR). Critically, however, EC effects differ from other instances of AARR in that the distal learning history that gives rise to AARR is brought to bear on one particular proximal regularity in the environment - namely - the pairing of stimuli. It might well be that other types of evaluative learning, such as those involved in mere exposure or approach/avoidance, also qualify as instances of AARR. However, in these cases, the learning history that gives rise to AARR is brought to bear on other proximal regularities (e.g., the repeated presentation of a stimulus or the relation between behavior and its consequences) which in turn leads to changes in liking.

The claim that a particular EC effect is an instance of AARR implies that this EC effect has core functional properties in common with other instances of AARR. More specifically, it implies that (a) the EC effect depends on the same distal learning history that allows for other instances of AARR, (b) manipulations of relational and functional cues should have an impact on that effect, and (c) that changes in liking should coincide with other changes in behavior that are instances of AARR. The first implication is difficult (although not impossible) to test empirically because it requires a manipulation of learning experiences during childhood (e.g., Luciano, Gómez Becerra, & Rodríguez Valverde, 2007). The second implication can more readily be tested by manipulating the nature or presence of relational and functional cues. The third implication can also be tested. If the change in liking of the CS due to CS-US pairings is an instance of AARR, then the change in liking should be just one element of acting as if the CS and US are related in a certain way. Hence, given appropriate contextual cues, the change in liking should coincide with other changes in behavior that are indicative of AARR, such as the ability to pick the US from a range of stimuli when presented with the CS as a sample[5].

In what follows, we evaluate whether there are logical and empirical arguments for the claim that many EC (and related) effects are instances of AARR and highlight the various contributions that this perspective offers. We first examine EC effects that are based on the pairing of stimuli in the presence or absence of other stimuli. We then examine so-called EC effects that have been established or manipulated via instruction, inference, and observation. Doing so will demonstrate how topographically distinct behavioral effects may actually be instances of the same underlying functional phenomenon (AARR).

Contribution 1: AARR allows for a sophisticated functional description of EC effects. Throughout much of the past three decades EC researchers have conceptualized ‘pairings’ as referring to the direct experience of a CS and US in spatio-temporal contiguity. A subset of these studies have also examined the impact of relational ‘qualifiers’ such as words or sentences on pairings given that these qualifiers provide additional information about the nature of the CS-US relation. In this section, we first consider the possibility that EC effects observed in the latter type of studies qualify as instances of AARR. We then examine whether EC effects in studies without explicit relational cues also qualify as instances of AARR.

EC effects with explicit relational cues. As we outlined above, two different types of contextual cues can moderate AARR by either (a) indicating the way in which events are related (relational cues) or (b) determining the functions that transfer between related stimuli (functional cues). If EC effects are an instance of AARR then these cues should also

[5] The relation between EC and AARR should now be clear: There might be instances of EC that are not instances of AARR (i.e., those instances in which stimulus pairings function as a mere proximal cause), there might be instances of EC that are instances of AARR (i.e., those instances in which stimulus pairings function as a proximal cue), and there are many instances of AARR that do not qualify as instances of EC (i.e., those instances that do not involve changes in liking or that involve other proximal cues than stimulus pairings).

moderate those EC effects. This idea provides a functional perspective on recent studies about the impact of relational information on EC. To illustrate, take the work of Förderer and Unkelbach (2012) who found that presenting CS and positively valenced US images in the presence of a relational cue "loves" resulted in a standard EC effect. However, when those same stimuli were related in the presence of a different cue ("loathes") the effect was completely reversed (see also Förderer & Unkelbach, 2011; Peters & Gawronski, 2011; Zanon, De Houwer, Gast, & Smith, 2014).

Interpreting these "relational EC effects" as instances of AARR provides a novel perspective on these phenomena. According to RFT, words such as "*goes with*", "*loves*", "*hates*", and "*is opposite to*" may function as contextual cues which moderate the impact of CS-US pairings on liking, much like the relational cues we described in Part III moderate other instances of AARR. Contextual cues can have this moderating effect because of a protracted history of operant learning that begins in infancy and continues through the lifetime of the individual (see Hughes & Barnes-Holmes, in press-a).

EC effects in the absence of explicit relational cues. As we also noted above, words are not the only stimuli that can function as relational cues. Physical objects (e.g., red traffic light), signs (e.g., mathematical or musical notation), gestures (e.g., a raised eyebrow during a job interview), or any other stimulus can signal that people should respond to other stimuli as being related in a particular way. According to this perspective, even spatio-temporal contiguity might constitute a proximal (relational) cue, signaling that stimuli which are paired in space and time are related to one another (e.g., Leader, Barnes, & Smeets, 1996; Zanon et al., 2014). This certainly seems plausible given that contiguity is a diffuse and ubiquitous property of the environment that is present in the 'background' of much of everyday life: it is present between the flick of a switch and subsequent presence of a light or a loud bang and a broken window. It is also present when an unknown stimulus (novel brand product) appears with a positively or negatively valenced image, or a neutral face is paired with an evaluative statement. Stated more precisely, humans encounter a vast array of stimuli that co-occur across a multitude of contexts throughout their entire lives. While the stimuli involved in these various learning experiences constantly differ, the functional relation between stimuli is held constant (i.e., stimuli that co-occur tend to be similar in some regard). Given a sufficient number of exemplars, irrelevant factors such as what the stimuli look, smell, taste or feel like may be “washed out” and the functional relation itself “abstracted” so that it becomes a proximal (contextual) cue for relating stimuli in the future. When participants encounter novel stimuli under similar environmental conditions (as they do in an EC experiment), spatio-temporal contiguity may be used as it has in the past - as a relational cue indicating that stimuli are similar or equivalent to one another. Once stimuli are related in this way their evaluative properties may change in line with the nature of the relation: in this case a transfer of evaluative functions from the US to the CS.

This functional analysis can be tested empirically. For example, if spatio-temporal contiguity is a proximal (relational) cue as we suggest, then manipulating its relational function should moderate the impact that stimulus pairings have on liking. Within the context of an experiment, it may be possible to establish contiguity as a cue meaning that stimuli should be related not as equivalent but as ‘opposite’ to one another. This could be achieved by intermixing the CS-US pairs with a large number of multiple filler trials on which two stimuli with an opposite meaning are presented (e.g., pictures of night and day, hot and cold, black and white stimuli). Following such training, the CS might acquire a valence opposite to the valence of the US with which it was paired given that previously contiguous stimuli were opposite to one another (see Zanon et al., 2014, Experiment 3, for an alternative strategy to assess the impact of spatio-temporal contiguity as a relational cue).

In short, we argue that contextual cues present in the proximal environment specify the manner in which stimuli are related and that these cues can either be obvious (e.g., words or symbols) or subtle (e.g., spatio-temporal contiguity). Even the pairing of stimuli in space and time can function as a proximal cue signaling that those events are related in a particular manner. For humans with a history of AARRing, the 'default' option may be to treat objects and events as similar to one another whenever they are simply presented together in space and time (e.g., Smyth et al., 2006; Leader & Barnes-Holmes, 2001). Yet when the relational properties of pairings as a cue are changed, or when people are exposed to other relational cues specifying alternative relationships between stimuli more complex relations and evaluative responses tend to emerge. The core message here is that humans may not only give meaning to, or semantically interpret the individual stimuli that are paired in EC studies, but also give meaning to or interpret the pairings themselves. EC effects are not solely determined by stimulus pairings but the way in which people *respond relationally* to those pairings.

Instructed, observed, and inference based EC effects as instances of AARR. EC effects are typically defined as changes in liking due to the pairing of stimuli. A number of researchers have sought to pair a CS with a US, not via repeated co-occurrence, but rather via instructions, inferences and observation. Once again, we believe that these learning effects also represent instances of AARR. For the time being, we will not address the question of whether these effects qualify as "genuine" instances of EC. This question will be discussed in detail below. For now, it suffices to say that some have argued that EC can be brought about via instructions, inferences, and observation. We now discuss how these phenomena relate to more traditional instances of EC from the perspective of RFT.

EC via instructions. According to RFT, instructions represent complex stimulus relations that serve to modify the psychological properties of stimuli in those relations (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004; Törneke, Luciano, & Valdivia Salas, 2008). Put simply, instructions are comprised of arbitrary stimuli (words) that typically 'stand for' other stimuli in the environment. For instance, the statement '*John helps his landlady with her trash*' does not involve the simple physical pairing of John with trashbags or his landlady. Rather it involves two types of stimulus relating. On the one hand, the words '*John*' and '*landlady*' have previously been related as equivalent to John and his landlady, respectively. On the other hand, these words are related via contextual cues contained in the statement itself (in this instance words such as '*helps*' and '*with* which lead to stimuli acquiring new or changing their existing psychological properties (in this case, it is John who becomes the helper and the Landlady who becomes the person who is helped). As an empirical example, consider, a recent study by Gast and De Houwer (2013) who observed a change in liking towards a CS (toothpaste) when participants were provided with the following instruction: "*If you see an image of toothpaste then a positive photo will appear*". For participants with a history of AARRing the word 'toothpaste' is typically treated as equivalent to actual toothpaste while the word 'positive photo' is treated as equivalent to previously encountered positive stimuli. At the same time, temporal and causal cues such as 'If' and 'Then' specify the order of events and their contiguous relationship to one another (i.e., that positive images only follow those of toothpaste and not other stimuli). By specifying this spatio-temporal relation the aforementioned cues lead to a transfer of evaluative properties from one stimulus (US) to another (CS) (see also De Houwer, 2006; Gast & De Houwer, 2013; Raes, De Houwer, De Schryver, Brass, & Kalisch, 2014; Zanon et al., 2014).

In other words, instruction-following represents a complex type of AARRing that can lead to rapid changes in existing and novel preferences for objects and events. The relational cues that comprise instructions are able to specify that stimuli were related in the past (e.g., 'CS caused cancer'), are currently related in the present ('CS is causing cancer'), or will be related at some point in the future ('CS will cause cancer') without the need for people to directly experience those events or even encounter the stimuli that they refer to. It should be noted that RFT also views changes in liking due to other types of verbal information, like statements ("*Mike cheated during a poker game*"; Peters & Gawronski, 2011; Mann & Ferguson, 2015), stories ("*Luppites are savage, ruthless and brutal*"; Gregg et al., 2006), relational qualifers ("*CS loathes cute kittens*"; Förderer & Unkelbach, 2012), and persuasive messages ("*Jonathan is a Yale-educated chemist who works for his state's Department of Water Resources*"; Smith, De Houwer, & Nosek, 2013) as instances of AARRing as well. In each case, arbitrary stimuli (words) are related to one another via relational cues, and specific psychological functions are ocassioned by functional cues, such that the psychological properties of words in those relations (and the objects that they refer to) are modifed as a result (for a detailed RFT account of instruction or rule-following see O'Hora & Barnes-Holmes, 2004). As we noted previously, in defining EC as an instance of AARR we do not argue that all instances of EC are necessarily the same or somehow identical. Rather the aforementioned EC effects appear to differ in terms of the proximal regularity that gives rise to AARR. We unpack this issue in greater detail below.

EC via inferences. Recently, several researchers have argued that EC effects may emerge on the basis of inferences about (rather than direct experience with) CS-US pairings. For instance, Gast and De Houwer (2012) exposed participants to an EC procedure in which a positive picture (US_{pos}) was directly paired with a gray square containing the number 1 (CS1) while a negative picture (US_{neg}) was paired with a gray square containing the number 2 (CS2). Following training, participants were informed that a neutral image (CS3) was always hidden behind the first grey square while another neutral image (CS4) was hidden behind the second square. Results indicated that automatic and self-reported liking was more positive for

CS3 than for CS4 despite the fact that neither CS was ever related with any US during training or instructions.

These findings bear a striking similarity to the aforementioned research on AARR. In particular, they mirror the fact that when humans encounter a small set of directly trained relations between stimuli, they often act as if those stimuli are related in novel and untrained ways. In the above mentioned study, for example, the emergence of inferred EC effects towards CS3 and CS4 likely reflected, from an RFT perspective, the formation of two equivalence relations comprised of valenced stimuli, numbered squares and neutral images (i.e., US_{pos}→CS1→CS3 and US_{neg}→CS2→CS4). Specifically, direct experience with one relation (e.g., US_{pos}→CS1) followed by instructions about another (e.g., CS1→CS3) may have allowed participants to act as if those stimuli were related in untrained ways (e.g., US_{pos}→CS3 and US_{neg}→CS4). Once stimuli were related as equivalent to one another, given appropriate contextual cues, a transfer of (evaluative) functions from one stimulus (US_{pos}) to another (CS3) may have taken place, explaining the observed pattern of automatic and self-reported responding. Thus, from an RFT perspective, when researchers use terms such as "inferences", "reasoning", and "deduction" in humans they refer to the generative properties of AARRing – namely – the ability to relate stimuli and events to one another in untrained and predictable ways (for a more detailed treatment see Barnes-Holmes, Keane, Barnes-Holmes & Smeets, 2000; Dack, Reed, & McHugh, 2010; Dougher et al., 2007; Dymond et al., 2008; Hayes et al., 2001; Valdivia-Salas et al., 2013).

EC via observation. Finally, changes in liking due to the pairing of stimuli cannot only take place via spatio-temporal contiguity, verbal instructions or inference, but can also stem from observing other people interact with the world. Observational learning refers to the capacity to learn how stimuli are related to one another by observing others when they are in contact with the environment. This type of learning allows organisms to change their existing (or acquire novel) behaviors without the need to ever directly encounter the aversive or appetitive consequences of their behavior for themselves (see Bandura, 1965; Zentall, 2012). For example, exposing rhesus monkeys to a real or videotaped counterpart reacting with fear towards a snake causes the observer to respond fearfully in the presence of, and attempt to avoid contact with, that same snake (Mineka & Cook, 1988). Similarly, children who observe their mother reacting with distress when she places her hand in cold water subsequently display a lower pain threshold when they have to place their own hand in water (Goodman & McGrath, 2003).

When applied to EC, observational learning refers to changes in liking that are due to the organism observing another organism interacting with the environment. For instance, watching another individual consume a drink (CS) and seeing them facially express a like or dislike for the drink (US) can cause the observer to produce a corresponding evaluative response towards that beverage (e.g., Baeyens et al., 2001; Baeyens, Kaes, Eelen, & Silverans, 1996). Although non-humans can modify their behavior based on imitation, modeling or social learning (Zentall, 2012), AARR may influence how humans learn via observation in two important ways. On the one hand, observing events in the environment may alter not only how people respond to those events but also impact how they respond to other related stimuli. For instance, suppose that a human watches a model bite into a large, sour lemon. Thereafter, the model emits a number of negative evaluative responses (e.g., displays signs of physical discomfort), and as a result, the observer avoids tasting lemons. Now suppose that the observer learns (either through training or instruction) that a novel brand product (Ettalas) tastes similar to lemons while a second product (Gageleer) is opposite to Ettalas. People may automatically display and self-report a preference for Gageleer relative to Ettalas and even opt to physically approach and consume the former relative to the latter – despite having never encountered either stimulus in the past. In other words, observing others interact with regularities in the environment may not only alter the psychological properties of those stimuli but also modify the properties of other related stimuli as well. The manner in which these properties are established or altered will once again depend on the relational and functional cues that control how stimuli are related.

On the other hand, AARR may also influence how people respond to the observed event itself. Take the above example of someone biting a sour lemon. Imagine that negative facial expressions are related with sour tastes and sour tastes with positive outcomes (e.g., by telling people that "*items which taste sour may appear to be unpleasant but they are actually extremely healthy for you*" or "*If you eat the lemon now I will give you a large sum of money later on*"). Now suppose that an individual is exposed to the same scenario as before in which a model bites into a lemon and emits a negative reaction. Rather than attempting to avoid any contact with lemons they may respond positively towards that stimulus in a variety of ways.

Conclusion. In short, there are good logical and empirical arguments to assume that different effects that have been reported in the EC literature may qualify as instances of AARR. Growing evidence indicates that once humans have learned to AARR they *respond relationally* to stimuli that are paired together and that this relating is (a) under the control of contextual cues in the environment, and that (b) these cues determine the magnitude, direction and nature of the change in liking. Contextual cues can vary in their complexity, from simple words and symbols to relatively complex statements and instructions. They can either be encountered though the individual's interactions with the environment or from observing the behavior of others in similar situations. Once people have the ability to AARR any stimulus or event can come to function as a contextual cue – even spatio-temporal contiguity between paired stimuli. Thus defining different EC effects as instances of AARR helps to systematize research (i.e., it has heuristic value). It highlights possible functional similarities between topographically distinct outcomes as well as possible functional differences between EC effects and other evaluative learning effects. It highlights that effects due to contiguity, instructions, observations, and inferences may be different instances of the same underlying phenomenon (AARR).

Before we continue, several points are worth noting here. First, our analysis does not imply that all effects that have been referred to as EC effects are identical simply because they are all instances of AARR. As we mentioned above, any given instance of AARR may be distinguished based on the type of proximal regularity that generates a relevant class (or classes) of AARRing (e.g., the pairing of stimuli, verbal instructions about the pairing of stimuli, verbal instructions about the properties of stimuli etc.). Although all of these instances of AARR are assumed to share some basic properties (e.g., depend on the distal learning history that allows for AARR, be sensitive to the nature and presence of relational and functional cues, coincide with other changes of behavior that are instances of AARR), different types of AARR might also differ in non-trivial ways depending on the proximal regularity that gave rise to AARR. For instance, the changes in liking due to contiguous stimulus pairings may differ in important ways from instructions about pairings, observing others interacting with stimulus pairings or the pairing of stimuli via inferences. The nature of these differences is an empirical matter. Regardless of the outcome of this future research, it seems likely that different classes of AARRing may result through interaction with different types of proximal regularities in the environment. Second, approaching the EC literature from the perspective of RFT not only has heuristic value (insofar as it identifies commonalities and differences across topographically distinct effects) but also has conceptual value. More specifically, it allows for complex phenomenon such as learning via instructions, inferences and deductions to be defined in purely functional terms. We will return to this point below.

Contribution 2: The concept of AARR can contribute to the debate on 'genuine' EC. Contribution 1 highlights an important point: the current definition of EC allows for considerable degress of freedom when determining what actually qualifies as a stimulus pairing. As we have seen, several authors have interpreted 'pairings' strictly as the directly experienced contiguous pairings of 'actual' stimuli (Jones et al., 2009; Walther et al., 2011) whereas others have interpreted 'pairings' in a more liberal manner, allowing them to be embedded in instructions (De Houwer, 2006; Field, 2006), to emerge via inference (Gast & De Houwer, 2012) or to take place based on observations of other organisms interacting in and with the environment (Baeyens et al., 2001). Thus the extent to which an outcome currently qualifies as a "genuine" or "real" instance of EC is often a matter of debate (see De Houwer & Hughes, submitted, and Gast et al., 2012 for more on this issue).

We see two ways in which this debate could be resolved. First, EC researchers could restrict the study of EC to those situations in which stimulus pairings function as a mere proximal cause of likes and dislikes. This would require EC effects to meet two strict conditions: (a) that stimulus pairings are the regularity responsible for changes in liking and (b) that those pairings are a mere proximal cause of liking rather than a cue for relational responding. The first condition can easily be tested by eliminating or controlling for other potential proximal regularities (see De Houwer, 2007; De Houwer, 2011b). However, it seems relatively more difficult to determine whether stimulus pairings are functioning as a mere proximal cause of liking or a cue for relational responding. One possible way to test this would be to keep the spatio-temporal properties of pairings constant while varying the relational properties of pairings as a contextual cue (for an example see Zanon et al., 2012). Note that one cannot settle this issue simply by inspecting the nature of the procedure. Even procedures that merely involve the pairing of stimuli and that are devoid of words or sentences might produce changes in liking only because stimulus pairings function as a proximal cue for the relation between those stimuli.

A second (and more pragmatic) option would be to restrict the study of EC to situations in which stimulus pairings function as the proximal regularity responsible for changes in liking, regardless of whether those pairings function as a proximal cause of liking or cue for relational responding. This position allows for other (distal and proximal) regularities to influence the function of stimulus pairings while ensuring that pairings are the proximal regularity responsible for the observed change in liking. Put another way, in order for an effect to qualify as an instance of EC, responding to the CS and US has to be under the control of stimulus pairings even when the function or "meaning" of those pairings has been altered by other regularities or events in the environment (e.g., Zanon et al., 2014). Such an approach would disqualify instructed and many inferred effects as instances of EC, not because they are dependent on distal learning experiences, but because they involve more than an impact of stimulus pairings as a proximal event (for a more detailed treatment of this topic see De Houwer & Hughes, submitted)[6].

The take home message here is that the current definition of EC places few constraints on how researchers interpret the term pairings and this can lead to disagreement about whether an effect actually qualifies as an instance of EC or not. We propose that EC can be delineated from other evaluative learning phenomena in two ways, either restrictively as changes in liking that result from stimulus pairings functioning as a mere proximal cause of liking, or more liberally, as changes in liking that results from stimulus pairings, regardless of whether those pairings function as a mere proximal cause or as a proximal cue. We favor the

[6] On a side-note we are not suggesting that the effects of instructions about stimulus pairings are irrelevant for EC researchers. On the contrary. If EC research centers on the effects of stimulus pairings on liking, then instructions about stimulus pairings represent a unique subclass of instructions given that they refer to a proximal event at the core of EC research (i.e., stimulus pairings). Studying the similarities and differences between effects that result from contiguous versus instructed stimulus pairings could unlock important information about the (mental) processes underlying both of these effects. It could also offer unique insight into the added value of experiencing events versus being instructed about those events (see Raes et al., 2014).

latter position because it remains true to the longstanding idea that EC is unique in its emphasis on stimulus pairings as a cause of liking (De Houwer, 2007) while acknowledging the potential complexity in the determinants of EC.

Although the argument that we developed in the previous paragraphs can be developed independent of the literature on RFT and AARR, we believe that this literature supports the argument. First, it suggests that any attempt to restrict EC to situations in which stimulus pairings function as a proximal cause of liking is going to encounter immediate problems. Based on research spanning more than 40 years, it appears that people do not respond based simply on the spatio-temporal properties of stimuli that they encounter (i.e., stimuli as they are present in the here and now) but rather based on how they relate those stimuli in the presence of contextual cues. The concept of AARR also allows for the possibility that there are two different types of EC effects: those that are instances of AARR and those that are not. For example, it may be that EC effects in humans who have acquired the ability to AARR are in fact instances of AARR, but EC effects observed in pre-verbal infants, severely verbally-impaired humans, or non-humans reflect changes in behavior due to the stimulus pairings functioning as a mere proximal cause of liking. Thus approaching EC as an instance of AARR directs attention away from questions about 'real' or 'genuine' effects and towards questions about the function of stimulus pairings and their role in changing liking. This in turn allows for a new cognitive perspective wherein EC is mediated by fundamentally different underlying mental processes depending upon an individual's level of verbal ability (for more see De Houwer & Hughes, submitted).

Second, the concept of AARR also offers insight into the learning experiences necessary to transform pairings from a mere proximal cause of liking to a proximal cue for relational responding. Existing knowledge concerning AARR may further refine our understanding of EC by specifying those distal regularities which are a prerequiste for EC effects involving words, sounds and other symbolic stimuli. Although researchers are of course free to appeal to other distal regularities that give rise to language or relational learning, and explore the impact of these regularities on stimulus pairings, the literature on AARR clearly articulates what these distal regularities look like and how they may be brought to bear on proximal events (see Hughes & Barnes-Holmes, in press-a,b).

Contribution 3: The concept of AARR can contribute to our understanding of human vs. non-human EC. If changes in liking can result from stimulus pairings functioning as a proximal cue, and if those learning experiences responsible for AARR are critical for establishing stimulus pairings as a proximal cue, then the concept of AARR may also inform the debate on human vs. non-human EC. This is because over forty years of research indicates that the ability to AARR is unique to, or at least highly elaborated in, humans while largely absence elsewhere in the animal kingdom (e.g., Dugdale & Lowe, 2000; Lionello-DeNolf, 2009). This failure to observe evidence of AARRing in non-humans was somewhat unexpected given the sheer number of learning principles that are common to humans and non-humans (e.g., reinforcement, punishment, generalization, discrimination, extinction, recovery and habituation). It was originally assumed that these principles would also stretch all the way up to complex phenomena such as language and higher-order cognition; yet a number of important findings emerged hinting at learning processes or principles that may be unique to, or at least largely elaborated in, some species relative to others (Hughes & Barnes-Holmes, 2014; Hayes, 1989). Today, and despite extensive efforts to find or train it, evidence for AARRing in nonhuman animals is still extremely limited. Although there are some indications that the most basic aspects of AARR might be found in nonhuman animals under strict laboratory conditions (e.g., Urcuioli, 2008; Zentall, Wasserman, & Urcuioli, 2014), the flexibility and complexity of AARRing as observed in humans far exceeds that seen in their nonhuman counterparts.[7]

These findings have important implications for our understanding of human and nonhuman EC. They suggest that there should be a clear divide between the types of EC effects observed in humans and those seen elsewhere in the animal kingdom. First, it seems reasonable to assume that the evaluative responses of animals can be changed whenever pairings between a CS and US function as a mere proximal cause of liking. For instance, pairing a tone (CS) and a tasty treat (US) in the presence of a dog may increase the probability that he will salivate and demonstrate other signs of "evaluative responding" (e.g., tail wagging, barking or jumping) whenever the tone is presented in the future. Second, changes in liking that result from pairing stimuli that have previously acquired their "meaning" via a history of AARRing may be beyond the reach of non-human organisms. Indeed, it seems almost obvious that pairing a tone (CS) with positive adjectives (US) will not produce a change in liking for non-humans in the same way that it does for humans. This is because the history of learning necessary to establish the "meaning" or function of many verbal stimuli is absent in one case and present in the other. Third, and more importantly for the present purposes, changes in liking that result from stimulus pairings functioning as a proximal cue signaling an arbitrarily applicable relational response will likely also be absent in non-humans and yet present in humans. As we have argued before, this is because pairings have different functions for organisms with than for organisms without the learning history which gives rise to AARR (i.e., pairings as a mere proximal cause vs. contextual cue). In

[7] It is worth noting that RFT has always remained agnostic to the possibility that AARR is a uniquely human capacity and it has never been argued that derived stimulus relating is forever beyond the grasp other species. Rather, RFT has simply viewed this claim as an empirical rather than purely theoretical one (see Dymond, Roche, & Barnes-Holmes, 2003). As the evidence currently stands, it seems likely that there is some "glass ceiling" in terms of relational complexity, contextual control, and generalizability that humans are capable of that is not evident elsewhere in the animal kingdom. That said, RFT allows for the possibility that providing nonhuman animals with the learning history that enables humans to AARR may also cause other organisms to behave in similar ways. Whether this potential can be realized is an empirical question.

short, pairing stimuli that have acquired their function or "meaning" via AARR should lead to EC effects that are evident in some species but not others. Likewise, when pairings function as a contextual cue for AARRing (or at the mental level: when pairings are interpreted or given semantic meaning), EC effects should be evident in certain species but not others, or it might have fundamentally different properties in verbal organisms (for which parings can function as a relational cue for AARRing) than in non-humans (for which pairings cannot function as a relational cue for AARRing). It may be that testing whether a behavior shows the properties of AARR provides one means in which to distinguish what types of EC effects different species can and cannot produce (for more on this topic see De Houwer & Hughes, submitted).

Contribution 4: The concept of AARR can help cognitive researchers refine existing and develop new cognitive theories of EC. The findings arising from research on AARR may also be used to constrain hypotheses and theories about the mental constructs that are assumed to mediate EC effects. For instance, some changes in liking that are instances of AARR appear to conflict with associative accounts, wherein the direct pairing of stimuli leads to the formation of unqualified links between mental representations in memory (Baeyens et al., 1992; Martin & Levey, 1994). In principle, from the perspective of typical associative accounts, presenting one stimulus with positive and another with negative images should result in the same mental associations and thus changes in liking regardless of the presence of relational cues. But research indicates that relational qualifiers do appear to moderate EC effects, both during (Förderer & Unkelbach, 2012) and after pairings take place (Peters & Gawronski, 2011; Zanon et al., 2014). Perhaps more importantly, they often lead to entirely different evaluative responses than those implied by the valence of the US that a CS was previously paired with. These accounts also encounter difficulties in explaining why EC effects emerge in the absence of any direct pairings between stimuli or why people will act as if stimuli are related in novel and untrained ways once a small set of relations have been directly trained.

Data from research on AARR may not only conflict with associative mental models but could be seen as providing support for propositional accounts that involve qualified links between mental representations in memory. The idea that EC effects result from the relating of stimuli under the control of contextual cues fits well with the idea that EC is mediated by the formation of propositions concerning those relations. Whereas associations simply convey the strength with which representations are linked in memory, propositions specify their strength, structure and content, and as such appear to articulate more readily with the RFT view of human cognition as being inherently relational. Likewise, while associations gradually develop with many experienced pairings (Dickinson, 2012; Gawronski & Bodenhausen, 2011) propositions can be formed on the basis of direct training or inferred via deductive reasoning and language (De Houwer, 2009). When combined, these two properties of propositions may explain the different patterns of evaluative responding that emerges when stimulus relations are established between and among stimuli.

For instance, imagine that people act as if two stimuli (a CS and US) are the same after observing them being paired. It may be that this behavior is mediated by a mental propositional representation such as '*CS is the same as US*'. If the relational "meaning" or function of pairings is altered so that contiguity now signals that stimuli are opposite to one another, and people act as if the CS and US are opposite, then this behavior may be mediated by another proposition - namely - that the "CS *is opposite to US"*. Indeed, any relation established between stimuli via contiguity, whether comparative ('CS-More than-US), hierarchical ('CS-Part of-US'), temporal ('CS-Comes before-US'), causal ('CS-Causes-US') or deictic ('I-am-US') could potentially lead to outcomes that are mediated by comparable propositions. At the same time, it is possible that the entirely untrained and novel ways in which people act towards stimuli after a small set of stimulus pairings may be mediated by participants making what we label a 'propositional leap'. In other words, learning that CS1 is the same as CS2 and that CS2 is the same as CS3, and then subsequently acting as if CS1 is the same as CS3 and CS3 is the same as CS1, may be mediated by the formation of a small set of propositions (e.g., '*CS1 is the same as CS2'* and '*CS2 is the same as CS3'*) which in turn generate an additional set of inferred propositions ('*CS1 is the same as CS3'*). It may be that these inferred propositions - rather than those that arise from direct experience - play a key role in changes in liking towards stimuli that were never paired with valenced objects or events in the past[8].

The core message here is twofold. First, the literature on RFT, and AARR more generally, represents a largely untapped reservoir of behavioral effects and procedures that cognitive researchers can use to improve existing or develop new mental theories. Just as defining a temper tantrum as an instance of reinforcement increases our ability to predict and influence such behavior (e.g., by specifying the antecedents and consequences that control such behavior), so too does defining EC as an instance of AARR. Indeed, when defined in this way, existing knowledge about the origins of AARR, its functional properties and boundary conditions can be used to further increase our understanding of EC effects. For instance, work in the RFT literature has already shown that pairings can function as a contextual cue for relational responding, and as a result, change how people act towards those stimuli (e.g., Leader & Barnes-Holmes, 2001). These studies typically exposed adults or children to a number of stimuli which are paired contiguously in space and time (e.g., A1 is

---

[8] It may be tempting to treat behavioral (AARR) and mental (propositions) phenomena as equivalent to one another in light of their apparent overlap (i.e., their flexibility, generativity and symbolic nature). However, as we pointed out above, it is important to realize that AARR refers to a pattern of contextually controlled behavior while propositions represent one possible mental mediator of the behavioral observations that functional researchers attempt to explain using the concept of AARR. In other words, AARR is not another name for propositional processes nor are propositions instances of AARR. For cognitive researchers, the former simply refers to a broad set of behavioral observations that need to be explained while the latter is one (of many) possible mental explanations for those behaviors.

paired with B1 and B1 then is paired with C1). Results consistently reveal that contiguity can function as a proximal cue signaling that stimuli are equivalent to one another (e.g., following such pairings participants tend to select A1 in the presence of C1; see Clayton & Hayes, 2004; Smyth et al., 2006). Extensive work has also examined the formation and manipulation of contextual cues and how those cues can lead to contrastive, comparative, opposition, and other types of relations between stimuli (for a review see Hughes & Barnes-Holmes, in press-a). This work could provide valuable insight into ongoing debates about the nature or "meaning" of stimulus pairings (proximal cause vs. cue) as well as the impact of relational qualifiers on stimulus pairings (e.g., Bar-Anan & Dahan, 2013; Moran & Bar-Anan, 2013; Förderer & Unkelbach, 2012, Zanon et al., 2014). Similarly, a wealth of data on AARR may also inform research on the generalization of EC effects (Valdivia-Salas et al., 2013), its sensitivity to forward/backward conditioning procedures (Barnes-Holmes et al., 2000), and resistance to extinction (Roche et al., 2008). Take, for example, the work of Valdivia-Salas and colleagues who found that when stimuli are related as equivalent to one another the EC effect established for one stimulus can "symbolically" generalize to other related stimuli even when those stimuli share no perceptual similarities with one another (note that stimulus generalization is typically restricted to perceptually similar stimuli). This may offer a new functional perspective on cognitive phenomena such as the "spreading attitude effect" (Walther, 2002) or "lateral attitude change" (Glaser et al., 2015), which often involve the generalization of evaluative responses along non-perceptual dimensions in ways that are difficult to explain using existing concepts such as second-order or sensory pre-conditioning.

Research that has been conducted in CBS may also lead to new predictions and ideas about EC. To illustrate, imagine that a positively valenced face (US) is repeatedly paired with a neutral face (CS) in the presence of the word "opposite". If we are correct and contiguity functions as a proximal cue signalling that stimuli are equivalent to one another then contiguity provides a source of information about the CS-US relation that is in direct contradiction to the relation implied by “opposite” (i.e. there is a competition between two different proximal cues). This may lead to a reduced rather than fully reversed EC effect relative to situations in which contiguity and the word “opposite” have the same relational meaning (i.e., a consistency or coherence between proximal cues). This competition-coherence hypothesis may provide a (functional) perspective on studies which have already pitted contiguity against other relational cues, either at the same time (Moran, Bar-Anan, & Nosek, 2015) or after a delay (Peters & Gawronski, 2011), as well as offer new predictions about the impact of relational qualifiers on liking. For instance, and as we alluded to above, it may be that stronger EC effects will emerge in those situations where contiguity and other cues signal the same relation between the CS and US compared to when they signal divergent or contradictory relations. Procedures developed within the functional tradition - such as the Matching to Sample task - could prove useful when asking and answering these questions. They would allow EC researchers to manipulate the meaning of proximal cues, establish relations between stimuli in ways that are under tight experimental control, and test whether EC effects co-occur with other instances of AARR.

The second core message is that given that the functional level is focused solely on the interplay between the environment and behavior, it does not place any *a priori* restrictions on what constructs mediate interactions between individuals and the environment. Actually, it does the opposite: it unlocks greater “theoretical freedom” such that researchers can deploy associations, propositions or any other mental concept in order to explain how changes in liking that could be seen as involving AARR come about. This may accelerate cognitive theorizing by freeing it from historical or pre-existing biases such as the notion that EC is exclusively associative in nature (for related arguments see Hughes, Barnes-Holmes, & De Houwer, 2011). However, and as noted in the introduction, findings from the functional literature do constrain mental theorizing in an *a posteriori* fashion. That is, they provide outcomes that mental models should be able to explain.

Contribution 5: The concept of AARR provides EC researchers with a means to functionally define complex psychological phenomena. Finally, the concept of AARR provides cognitive researchers with a way of talking about an entire spectrum of EC and related effects in purely functional terms. Given that there are always multiple possible mental theories of behavior (and because behavior is never determined by just one set of mental processes), it is important to maintain a clear conceptual distinction between the to-be-explained behavioral effect and the proposed mental theory (see De Houwer & Moors, 2015). Over the past decade this conceptual separation has been widely adopted within the EC literature. However, the concept of AARR allows one to also define related phenomena such as EC via instructions, inferences, and observation in functional terms. Because knowledge about these effects can be useful for EC research regardless of whether they are considered to be “real” EC effects, it is also good for EC researchers to be able to conceptualize these effects in functional terms. Although effects due to instructions, inferences, and observation have historically been defined at the mental level of analysis, there is no *a priori* reason why they cannot be defined in purely functional terms. Whereas the functional literature may have been ill-equipped to grapple with these concepts in the past, times have changed. With the development of RFT, and its emphasis on the concept of AARR, many complex psychological phenomena (including instructed and inferential learning) can and have been conceptualized, studied and explored at the functional level of analysis (see Hayes et al., 2001; Hughes & Barnes-Holmes, in press-b).

Adopting a sophisticated functional (analytic-abstractive) language for EC related effects may allow for even greater theoretical innovation at the mental level, insofar as it ensures that no pre-existing restrictions are placed on the potential mental processes and representations that can mediate the effect of stimulus pairings on liking. For instance, it may be that instructed, inferred, and observed EC effects are mediated by some complex interplay between mental associations in memory. Although such an explanation of the behavioral observations that AARR attempts to explain (at the functional level) appears difficult to reconcile with the literature, it is not excluded from consideration on an *a priori* basis. In short, CBS may help to equip cognitive researchers with a sophisticated non-mental language that maintains a conceptual distinction between the event that needs to be explained (behavioral effect) and the event that is used to explain (mental mediator).

Conclusion

The current paper introduced an intellectual tradition known as CBS, a functional account of human language and cognition known as RFT, as well as a behavioral phenomenon known as AARR. We then forwarded a novel perspective on EC wherein stimulus pairings function as either a mere proximal cause of liking or a proximal cue signaling that the CS and US are related. This latter possibility leads to the idea that the impact of proximal regularities (stimulus pairings) on liking may be moderated by distal regularities in the organism's past. While fully recognizing that other accounts are possible, we draw upon developments in the RFT and EC literatures to propose that a specific set of distal regularities (those that give rise to AARRing) may advance our understanding of EC. According to this perspective, humans not only give meaning to the individual stimuli that are paired in EC studies but also to pairings as a proximal event. That is, they *respond relationally* to the fact that stimuli are paired together and this behavior is (a) under the control of contextual cues in the environment, and (b) these cues determine the magnitude, direction and nature of the change in liking. Not only instructions, observations and inferences but also contiguous stimulus pairings lead to changes in liking because of a host of learning experiences that have endowed the individual with the ability to AARR. In this way, EC effects that have been traditionally attributed to the presence of a single (proximal) regularity may in fact be moderated by other (distal) regularities. Indeed, if one adopts an RFT perspective, they may even represent different instances of the same behavioral phenomenon (AARR).

We presented logical and empirical arguments for the usefulness of this conceptualization and discussed how it contributes to our understanding of EC in a variety of ways, from heuristically organizing existing and predicting novel effects, contributing to debates on "genuine" and human vs. non-human EC effects, as well as facilitating the development and refinement of cognitive theories of EC. We also showed how the functional analytic concept of AARR provides a means to discuss and define increasingly complex phenomena such as so-called EC via instructions and inferences in purely functional terms, and in doing so, enables researchers to maintain a firm separation between mental and functional levels of analysis.

We can see three ways in which cognitively-oriented EC researchers might respond to our arguments. First, they might take note of the idea that some instances of EC go beyond the effect of stimulus pairings as a mere proximal cause and that RFT might help us understand those instances but dismiss these instances as lying beyond the scope of EC research. Such a position would imply that EC research should focus on changes in liking in which stimulus pairings are a mere proximal cause. Second, they could include instances of EC in which stimulus pairings are used to give meaning to stimulus relations but without taking on board the CBS literature. Third, they could turn to the CBS literature as a source of inspiration for their cognitive analysis of EC, including instances of EC in which stimulus pairings are used as a proximal cue for giving meaning to stimulus relations. Based on the arguments presented above, we believe that the third option would optimize progress in our understanding of EC. In any case, our hope is that this paper serves to stimulate new debate on the role that AARR in particular, and distal regularities in general, may play in EC and related phenomena, and that it leads to renewed dialogue between cognitive and functional researchers interested in the study of likes and dislikes.

References

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation, 27,* 200-241.

Baeyens, F., Eelen, P., Crombez, G., & De Houwer, J. (2001). On the role of beliefs in observational flavor conditioning. *Current Psychology, 20,* 183-203.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning; acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behavior Research and Therapy, 30,* 133–142.

Balas, R., & Gawronski, B. (2012). On the intentional control of conditioned evaluative responses. *Learning and Motivation*, *43*, 89–98.

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology, 1,* 589-595.

Bar-Anan, Y., & Dahan, N. (2013). The effect of comparative context on evaluative conditioning. *Cognition & emotion*, *27(2),* 367-375.

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63,* 2313-2335.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.

Barnes-Holmes, D., & Hussey, I. (2015). The functional-cognitive meta-theoretical framework: Reflections, possible clarifications and how to move forward. *International Journal of Psychology*. DOI: 10.1002/ijop.12166

Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. M. (2000). A derived transformation of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record, 50,* 493-511.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., & Luciano, C. (2004). The derived transfer of mood functions through equivalence relations. *The Psychological Record, 54,* 95-114.

Bechtel, W. (2008). Mechanisms in cognitive psychology: What are the operations? *Philosophy of Science, 75,* 995-1007.

Boakes, R. A., Albertella, L., & Harris, J. A. (2007). Expression of flavor preference depends on type of test and on recent drinking history. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 327-338.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968 -1987, *Psychological Bulletin, 106,* 263–289.

Cahill, J., Barnes-Holmes, Y., Barnes-Holmes, D., Rodríguez-Valverde, M., Luciano, C., & Smeets, P. M. (2007). The derived transfer and reversal of mood functions through equivalence relations II. *The Psychological Record, 57(3),* 373-389.

Catania, A. C. (1997). *Learning*. Upper Saddle River, NJ: Prentice-Hall.

Chiesa, M. (1992). Radical behaviorism and scientific frameworks: From mechanistic to relational accounts. *American Psychologist, 47,* 1287-1299.

Clayton, M. C., & Hayes, L. J. (2004). A comparison of match-to-sample and respondent -type training of equivalence classes. *The Psychological Record, 54(4),* 579-602.

Corneille, O., Yzerbyt, V., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology, 45,* 279–282.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37,* 176–187.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230–241.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37,* 1–20.

De Houwer, J. (2011a). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202-209.

De Houwer, J. (2011b). Evaluative conditioning: Methodological considerations. In Klauer, K.C., Stahl, C., & Voss, A. (Eds.) *Cognitive methods in social psychology* (pp. 124 -147). New York: Guilford.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20(4),* 631-642. doi: 10.3758/s13423-013-0386-3

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, 24(1), 252 -287.

De Houwer, J., & Hughes, S. (submitted). Why is evaluative conditioning important? On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion.

De Houwer, J., & Moors, A. (2015). Levels of analysis in social psychology. In B. Gawronski & G. Bodenhausen (Eds.), *Theory and Explanation in Social Psychology* (pp. 24-40). NY: Guilford.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135,* 347–368.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853–869.

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society of Biological Sciences, 367,* 2733-2742.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345-355.

Dougher, M. J., Hamilton, D., Fink, B., & 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.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5–22.

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

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2008). Derived avoidance learning: Transformation of avoidance response functions in accordance with same and opposite relational frames. *The Psychological Record*, *58,* 269-286.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69,* 1013–1027.

Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology*, *44*(7), 657 -670.

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes, *Cognition & Emotion, 25(4)*, 639-656.

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS–US relations, *Cognition & Emotion, 26(3),* 534-540.

Field, A. P. (2006). I don't like it because it eats sprouts: Conditioning preferences in children. *Behavior Research and Therapy, 44,* 439-455.

Fulcher, E. P., Mathews, A., & Hammerl, M. (2008). Rapid acquisition of emotional information and attentional bias in anxious children. *Journal of Behavior Therapy and Experimental Psychiatry, 39,* 321-339.

Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science, 321,* 1100-1102.

Gannon, S., Roche, B., Kanter, J., Forsyth, J. P., & Linehan, C. (2011). A derived relations analysis of approach-avoidance conflict: Implications for the behavioral analysis of human anxiety. *The Psychological Record, 61,* 227-252.

Gardner, H. (1987). The mind's new science: A history of the cognitive revolution. New York: Basic Books.

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli, *The Quarterly Journal of Experimental Psychology, 65(9),* 1657-1674.

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions. *Learning and Motivation, 43,* 79-88.

Gast, A & De Houwer, J. (2013). The influence of extinction and counterconditioning instructions on evaluative conditioning effects. *Learning and Motivation*, Advance online publication. doi: http://dx.doi.org/10.1016/j.lmot.2013.03.003

Gast, A., & Rothermund, K. (2011). I like it because I said that I like it. Evaluative conditioning effects can be based on stimulus–response learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37*, 466–476.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General, 139*, 683–701.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59-127.

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology, 48*, 617-623.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research, 35*, 178 –188.

Gil, E., Luciano, C., Ruiz, F. J., & Valdivia-Salas, V. (2012). A Preliminary Demonstration of Transformation of Functions through Hierarchical Relations. *International Journal of Psychology and Psychological Therapy, 12*, 1-19.

Giurfa, M., Zhang, S. W., Jennett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of sameness and difference in an insect. *Nature, 410*, 930-933.

Glaser, T., Dickel, N., Liersch, B., Rees, J., Sussenbach, P., Bohner, G. (2015). Lateral attitude change. *Personality and Social Psychology Review, 19(3)*, 257-276.

Goodman, J. E., & McGrath, P. J. (2003). Mothers' modeling influences children's pain during a cold-pressor task. *Pain, 104,* 559-565.

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

Gregg, A. I., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90,* 1–20.

Hammerl, M., & Grabitz, H. J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning and Motivation, 27,* 278–293.

Hammerl, M., & Grabitz, H. J. (2000). Affective-evaluative learning in humans: A form of associative learning or only an artifact? *Learning and Motivation, 31,* 345–363.

Harmon, K., Strong, R., & Pasnak, R. (1982). Relational responses in tests of transposition with rhesus monkeys. *Learning and Motivation, 13(4),* 495-504.

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51,* 385–392.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1,* 1-16.

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

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy, 44(2),* 180-198.

Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology, 52,* 180–186.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136,* 390–421.

Hollands, G. J., Prestwich, A. & Marteau, T. M. (2011). Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychology, 30(2),*195-203.

Houben, K. , Schoenmakers, T. M., & Wiers, R. W. (2010). I didn't feel like drinking but I don't know why: The effects of evaluative conditioning on alcohol-related attitudes, craving and behavior. *Addictive Behaviors*, *35*, 1161-1163.

Hughes, S., & Barnes-Holmes, D. (in press-a). Relational Frame Theory: The Basic Account. In S. Hayes, D. Barnes-Holmes, R. Zettle, and T. Biglan (Eds.), *Handbook of Contextual Behavioral Science*. New York: Wiley-Blackwell.

Hughes, S., & Barnes-Holmes, D. (in press-b). Relational Frame Theory: Implications for the Study of Human Language and Cognition. In S. Hayes, D. Barnes-Holmes, R. Zettle, and T. Biglan (Eds.), *Handbook of Contextual Behavioral Science*. New York: Wiley -Blackwell.

Hughes, S., & Barnes-Holmes, D. (2014). Associative concept learning, stimulus equivalence, and Relational Frame Theory: Working out the similarities and differences between human and nonhuman behavior. *Journal of the experimental analysis of behavior, 101(1)*, 156-160.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The Dominance of Associative Theorizing in Implicit Attitude Research: Propositional and Behavioral Alternatives. *The Psychological Record, 61*, 465–498.

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

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141(3),* 539 –557.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning, *Journal of Personality and Social Psychology, 96*, 933–48.

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology, 58*, 31-38.

Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., & Stark, R. (2009). Neural,electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture–picture conditioning paradigm. *Neuroscience*, *158,* 721–731.

Leader, G., & Barnes-Holmes, D. (2001). Matching-to-sample and respondent-type training as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record, 51*, 429-444.

Leader, G., Barnes, D., & Smeets, P. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record, 46,* 685-706.

Leader, G., Barnes-Holmes, D., & Smeets, P. M. (2000). Establishing equivalence relations using a respondent-type procedure III. *The Psychological Record, 50,* 63-78.

Liefooghe, B., & De Houwer, J. (*in press*). A functional approach for research on cognitive control: Analyzing cognitive control tasks and their effects in terms of operant conditioning. *The International Journal of Psychology*.

Levin, M. E., & Hayes, S. C. (2009). ACT, RFT, and contextual behavioral science. In J. T. Blackledge, J. Ciarrochi, & F. P. Deane (Eds.), *Acceptance and Commitment Therapy: Contemporary research and practice* (pp. 1-40). Sydney: Australian Academic Press.

Lionello-DeNolf, K. M. (2009). The search for symmetry: 25 years in review. *Learning & Behavior, 37,* 188–203.

Lubow, R.E., & Gewirtz, J. C. (1995). Latent inhibition in humans: Data, theory, and implications for schizophrenia. *Psychological Bulletin, 117,* 87-103.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108(6),* 823-849.

Martin, I., & Levey, A. B. (1994). The evaluative response: Primitive but necessary. *Behavior Research and Therapy, 32,* 301–305.

McHugh, L., & Stewart, I. (2012). *The self and perspective taking: Contributions and applications from modern behavioral science*. Oakland: New Harbinger Publications.

Meiser, T. (2011). Much pain, little gain? Paradigm-specific models and methods in experimental psychology. *Perspectives on Psychological Science, 6,* 183-191.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T. R. Zentall, Ed. *Social learning: psychological and biological perspectives* (pp. 51–72). New York: Routledge.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,*183–198.

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion, 4,* 743-752.

Moran, T., Bar-Anan, Y., & Nosek, B. (2015). Processing goals moderate the effect of co -occurrence on automatic evaluation. *Journal of Experimental Social Psychology, 60,* 157-162.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15,* 152-159.

O'Hora, D., & Barnes-Holmes, D. (2004). Instructional control: Developing a relational frame analysis. *International Journal of Psychology and Psychological Therapy, 4,* 263-284.

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

O'Hora, D., Pelaez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record, 58,* 569-584.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413–417.

Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin, 37,* 557-569.

Pettigrew, T. F. & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory, *Journal of Personality and Social Psychology, 90,* 751–83.

Pinter, B., & Greenwald, A. G. (2004). Exploring implicit partisanship: Enigmatic (but genuine) group identification and attraction. *Group Processes & Intergroup Relations, 7(3),* 283–396.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item -based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33,* 130-144.

Proctor, R. W., & Urcuioli, P. J., (in press). Functional Relations and Cognitive Psychology: Lessons from Human Performance and Animal Research. *The International Journal of Psychology*.

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US Pairings Actually Matter? A Within-Subject Comparison of Instructed Fear Conditioning with and without Actual CS-US Pairings. *PLoS ONE 9(1),* e84888. doi:10.1371/journal.pone.0084888

Reese, H. W. (1968). *The perception of stimulus relations: Discrimination learning and transposition*. New York: Academic Press.

Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding: Applications for learners with autism and other developmental disabilities.* Oakland, CA: New Harbinger Publications, Inc.

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,* 267-292.

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*, 443-464.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91,* 995–1008.

Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record*, *50,* 339-354.

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin*, *39,* 193-205.

Smyth, S., Barnes-Holmes, D., & Forsyth, J. P. (2006). A derived transfer of simple discrimination and self-reported arousal functions in spider fearful and non-spider fearful participants. *Journal of the Experimental Analysis of Behavior, 85(2),* 223 -246.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes, 35(2),* 286–91.

Törneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical applications*. Oakland, CA: New Harbinger Publications, Inc.

Törneke, N., Luciano, C., & Valdivia Salas, S. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology and Psychological Therapy, 8(2),* 141-156.

Urcuioli, P. J. (2008). Associative symmetry, anti-symmetry, and a theory of pigeons' equivalence class formation. *Journal of the Experimental Analysis of Behavior, 90(3),* 257-282.

van Reekum, C. M., van den Berg, H., & Frijda, N. H. (1999). Cross-modal preference acquisition: evaluative conditioning of pictures by affective olfactory and auditory cues. *Cognition and Emotion, 13,* 831–836.

Valdivia-Salas, S., Dougher, M., & Luciano, C. (2013). Derived relations and generalized alteration of preferences. *Learning and Behavior*, *41*, 205-217.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison: Systematic empirical analyses. *The Psychological Record, 58,* 365-390.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82,* 919–934.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the backdoor: The US-revaluation effect. *Cognition and Emotion, 23,* 889 -917.

Walther, E., Langer, T., Weil, R., & Komischke, M. (2011). Preferences surf on the currents of words: Implicit verb causality influences evaluative conditioning. *European Journal of Social Psychology, 41,* 17–22.

Walther, E. & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 454–459.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19*, 175-196.

Whelan, R., & Barnes-Holmes, D. (2004). The transformation of consequential functions in accordance with the relational frames of same and opposite. *Journal of the Experimental Analysis of Behavior, 82,* 177-195.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship to questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.

Yamazaki, Y., Saiki, M., Inada, M., Iriki, A., & Watanabe, S. (2014). Transposition and its generalization in common marmosets. *Journal of Experimental Psychology: Animal Learning and Cognition, 40(3)*, 317.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation, 43,* 155–165.

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning?. *The Quarterly Journal of Experimental Psychology, 67(11),* 2105-2122.

Zentall, T.R. (2012). Perspectives on observational learning in animals. *Journal of Comparative Psychology, 126*, 114–128.

Zentall, T. R., Wasserman, E. A., & Urcuioli, P. J. (2014). Associative concept learning in animals. *Journal of the experimental analysis of behavior, 101(1),* 130-151.