

The National University of Ireland Maynooth



NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Why We Like What We Like: A Functional Approach to the Study of Human
Evaluative Responding

Thesis submitted to the Department of Psychology, Faculty of Science, in
fulfilment of the requirements for the degree of Doctor of Philosophy,
National University of Ireland, Maynooth.

Sean Hughes B.A. (Hons.)

October 2012

Head of Department: Dr. Fiona Lyddy

Research Supervisor: Professor Dermot Barnes-Holmes

Abstract

The current thesis set out to investigate whether the behavioural process of derived stimulus relating could facilitate a better understanding, prediction and influence of human likes and dislikes than that afforded by direct contingency accounts alone. Across a series of six experiments and in the absence of co-occurrence, reinforcement or instruction, stimuli spontaneously acquired evaluative functions by participating in derived coordination, opposition and comparative relations. By exerting fine-grained contextual control over how Pokémon characters, fictitious brand products or potential prizes were related to one another, we systematically manipulated the direction and magnitude of evaluative responding. These relational effects were evident when a range of direct and indirect (IAT, IRAP and Affective Priming) tasks were employed. When taken together, our work suggests that a sophisticated experimental analysis of evaluative responding cannot focus solely on simple stimulus pairings – at least where verbally trained humans are concerned. Rather, a comprehensive understanding of human likes and dislikes requires a shift in current research practices, with direct and derived stimulus relations explored in tandem.

Acknowledgements

I would like to acknowledge a number of friends and colleagues for their support, encouragement and guidance over the last four years. In particular, my special thanks go to:

Professor Dermot Barnes-Holmes

For introducing me to the academic sweet-shop and virtues of adhering to a strict diet of Mars-Bars. I am truly grateful for the intellectual freedom and unwavering support you have provided. I could not have asked for a better mentor, inspiration and (hopefully) friend

Professor Jan De Houwer

For your scientific rigour, invaluable insight and guidance when exploring novel intellectual islands.

My Family

Mum, Dad, Ciaran, Ashling, Rian and Meabh

For shaping me into the person I am today, I love you all

Corinna

For setting out on this journey with me and being there every step of the way

Ian

For providing critical debate and witty banter in equal measure

My Friends,

Evan, Pearse, Jamie, Anna and Conor

For Furious George to The Manhattan Syndrome and everything in-between

Table of Contents

Abstract	ii	
Acknowledgements	iii	
Table of Contents	iv	
List of Figures	vi	
List of Appendices	ix	
Chapter 1: Evaluation at the Mechanistic and Functional Levels of Analysis:		
A Review	1	
1.1 A Mechanistic Approach to Evaluative Responding.....	3	
1.2 Evaluative Conditioning at the Procedural Level.....	4	
1.3 Evaluative Conditioning at the Effect Level.....	6	
1.4 Evaluative Conditioning at the Mental Process Level.....	8	
1.5 Summary.....	12	
1.6 A Functional Approach to Psychological Science.....	14	
1.7 Relational Frame Theory.....	19	
1.8 Evaluative Responding as Relational Responding.....	23	
1.9 Respondent Preparations.....	23	
1.10 Non-Respondent Preparations.....	24	
1.11 Summary.....	26	
1.12 Overview of Current Research.....	29	
Chapter 2: A Transformation of Functions through Mutually Entailed Relations as Measured by the IAT and Self-Report Tasks		33
2.1 Experiment 1.....	34	
2.2 Method.....	36	
2.3 Results.....	45	
2.4 Discussion.....	50	
Chapter 3: A Derived Transformation of Functions through Combinatorially Entailed Relations as Measured by the IAT, IRAP and Self-Report Tasks		56
3.1 Experiment 2.....	56	
3.2 Method.....	59	
3.3 Results and Discussion.....	64	
3.4 Experiment 3.....	70	
3.5 Method.....	71	
3.6 Results and Discussion.....	74	
3.7 General Discussion.....	79	
Chapter 4: A Derived Transformation of Functions through Coordination and Opposition Relations as Measured by the IAT, Affective Priming and Self-Report Tasks		87
4.1 Experiment 4.....	87	
4.2 Method.....	90	
4.3 Results.....	97	
4.3 Discussion.....	105	

Chapter 5: A Derived Transformation of Functions through Comparative Relations as Measured by the IAT and Self-Report Tasks.....	112
5.1 Experiment 5.....	112
5.2 Method.....	115
5.3 Results and Discussion.....	122
5.4 Experiment 6.....	126
5.5 Method.....	127
5.6 Results and Discussion.....	128
5.7 General Discussion.....	132
Chapter 6 General Discussion.....	137
6.1 Overview of the Current Research Programme.....	137
6.2 A Functional Approach to Evaluative Responding.....	137
6.3 Summary of the Current Research.....	139
6.4 Conceptual Issues.....	143
6.5 Implications for the Mechanistic Approach to Evaluative Responding.....	151
6.6 Interplay Between the Functional and Mechanistic Traditions.....	156
6.7 Limitations and Future Directions.....	162
6.7 Conclusion.....	165
References.....	167
Appendices.....	190

List of Figures

Figure 2.1	Examples of the similarity and opposition training and testing trials. Each trial consisted of two pictures at the top of the screen and the two to-be-trained contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).....	39
Figure 2.2	Examples of the four relational training and testing trials involved in establishing either a coordination or opposition relation between a Pokémon (CS) and valenced image (US). Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).....	42
Figure 2.3	Mean likeability scores for each of the four Pokémon characters as a function of the valenced images and contextual cue it was paired with. Error bars represent standard errors.....	48
Figure 2.4	Mean D-IAT scores as a function of block order for both the coordination and opposition IATs. Error bars represent standard errors.....	50
Figure 3.1	Schematic representation of the two experimentally established relational networks. In each case, three arbitrary Pokémon characters were first related to one another using the ‘Same’ cue (<i>Pokémon1- Pokémon2- Pokémon3</i> and <i>Pokémon4-Pokémon5- Pokémon6</i>). Thereafter, an opposition relation was established between the first member of either relation and positive or negative images (<i>Pokémon1-Opposite-Positive</i> and <i>Pokémon 4-Opposite-Negative</i>).....	61
Figure 3.2	Examples of the four trials involved in the first phase of relational training. Each trial consisted of two Pokémon at the top of the screen and the two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).....	63
Figure 3.3	Mean likeability scores for each of the Pokémon characters as a function of derivation test performance (pass vs. fail). Error bars represent stand errors.....	67
Figure 3.4	Mean D-IAT scores as a function of derivation test performance (pass vs. fail) and derivation test time (before vs. after). Error bars represent standard errors.....	69

Figure 3.5	Examples of the four trial-types used in the IRAP. A label stimulus (Pokémon 3 or 6), target word (nasty, nice, disgusting, pleasant, etc.), and two response options (True and False) appeared simultaneously on each trial. Selecting the option deemed correct on any given trial resulted in “Correct” being displayed in the middle of the screen while selecting the option deemed incorrect caused “Incorrect” to appear.....73
Figure 3.6	Mean likeability scores for each of the Pokémon characters as a function of derivation test performance (pass vs. fail). Error bars represent standard errors.....76
Figure 3.7	Mean D_{IRAP} scores for each of the four trial-types as a function of derivation test performance (pass vs. fail). A positive score indicates responding in-accordance with prior training while a negative score indicates responses that are inconsistent with training. Error bars represent standard errors.....78
Figure 4.1	Schematic representation of the trained relations in Experiment 4. Two, three member coordination relations were established each comprised of three fictitious brand products (<i>Pardal-Same-Zatte-Same-Ettalas</i> and <i>Ciney-Same-Witkap-Same-Gageleer</i>). Half of the participants were then trained to relate <i>Pardal-Same-Positive</i> and <i>Ciney-Same-Negative</i> while the other half were trained to relate <i>Pardal-Opposite-Positive</i> and <i>Ciney-Opposite-Negative</i>92
Figure 4.2	Examples of the four trials involved in the first phase of relational training. Each trial displayed two brand names at the top of the screen and the two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).....93
Figure 4.3	Mean likeability scores for Pardal, Zatte and Ettalas as a function of relation (coordination vs. opposition) and derivation test performance (pass vs. fail). Error bars indicate standard errors.....99
Figure 4.4	Mean likeability scores for Ciney, Witkap and Gageleer as a function of relation (coordination vs. opposition) and derivation test performance (pass vs. fail). Error bars indicate standard errors.....101
Figure 4.5	Mean D_{IAT} scores as a function of derivation test performance (pass vs. fail) and the type of relation established (coordination vs. opposition). Error bars represent standard errors.....102
Figure 4.6	Mean response latencies for the positive and negative primes as a function of the type of relation established (coordination vs. opposition). Error bars represent standard errors. Note that the priming effect is obtained by comparing the response latencies for the positive prime to that of the negative prime.....104

Figure 5.1	Examples of the ‘More than’ and ‘Less than’ training and testing trials. Each trial consisted of two pictures at the top of the screen and two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).....117
Figure 5.2	Schematic representation of the comparative relation established in Experiment 5. Five fictitious brand products participated in the relation (<i>Pardal-LessThan-Zatte-LessThan-Ettalas-LessThan-Ciney-LessThan-Witkap</i>). A consequential function was then established for Pardal by pairing it with repeated access to €.01 and Zatte with €.25 per trial.....118
Figure 5.3	Examples of the four trials involved in the first phase of (comparative) relational training. Each trial displayed two prize names at the top of the screen and the two contextual cues at the bottom of the screen.....119
Figure 5.4	Mean likeability scores for Zatte, Ettalas and Ciney as a function of derivation test performance (pass vs. fail). Error bars indicate standard errors.....124
Figure 5.5	Mean D-IAT scores for Zatte relative to Ettalas, Ettalas relative to Ciney and Ciney relative to Zatte. A positive score indicates a response bias for the stimulus further along the comparative relation. Error bars represent standard errors.....125
Figure 5.6	Mean likeability scores for Zatte, Ettalas and Ciney. Error bars indicate standard errors.....130
Figure 5.7	Mean D-IAT scores for Zatte relative to Ettalas, Ettalas relative to Ciney and Ciney relative to Zatte. A positive score indicates a response bias for the stimulus further along the comparative relation. Error bars represent standard errors.....131

List of Appendices

Appendix A: Consent Form.....	190
Appendix B: Contextual Cue Training Instructions.....	192
Appendix C: Relational Training Instructions.....	193
Appendix D: Stimulus Rating Task.....	194
Appendix E: Contextual Cue Rating Task.....	195
Appendix F Demand Compliance Task.....	196

Chapter 1: Evaluation at the Mechanistic and Functional Levels of Analysis: A Review

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 (Watson & Rayner, 1920; Martin & Levey, 1978; De Houwer, 2007). These evaluative responses are thought to play a causal role in a diverse spectrum of psychological phenomena, including consumer choice behaviours (Gibson, 2008; Hollands, Prestwich & Marteau, 2011), in-group favouritism and stigmatization (Walther, Nagengast & Trasselli, 2005) as well as self-esteem (Dijksterhuis, 2004) and voting intentions (Galdi, Arcuri & Gawronski, 2008). In order to understand, predict and influence these behaviours in a sophisticated manner, the learning processes involved in the formation and change of evaluative responding must first be identified.

As in any area of (psychological) science, researchers interested in evaluative responding have explicitly or implicitly adopted a set of philosophical assumptions about the research domain, appropriate units of analysis and relevant truth criteria. These pre-analytic assumptions provide the philosophical scaffold upon which individual theories have been built, methodologies crafted and empirical findings interpreted. Although a number of philosophical frameworks or “world-views” are available to guide scientific activity (Pepper, 1942; Hayes, Hayes & Reese, 1988), most research in this area has been conducted by psychologists subscribing to either a cognitive (mechanistic) or functional (contextual) position (referred to hereafter as the mechanistic and functional approaches respectively). In the following chapter we present a detailed overview of the core assumptions and analytic strategies upon which the mechanistic and functional traditions have been built. As we shall see, both traditions have sought to understand, predict, and in some cases influence, evaluative responding using radically different conceptual, theoretical and methodological tools. The first section will open with a brief review of the mechanistic literature and examine

how the study of evaluation has been tackled at the procedure, effect and mental process levels of analysis. In particular, we focus our attention on a phenomenon known as evaluative conditioning (i.e., a change in liking that results from the pairing of stimuli) which is argued to represent an important avenue through which evaluative responses may be established or modified (De Houwer, 2011a; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). As we shall see, EC is often conceptualised as a form of associative or Pavlovian conditioning (PC), despite the fact that it demonstrates several response properties and operates under environmental conditions that are uncharacteristic of respondent learning. What will also become evident is that much of this work has been driven by a set of mechanistic assumptions that specify a change in liking (behaviour) that occurs due to the pairing of stimuli (environment) in terms of mediating mental processes and representations. The result is an empirical agenda focused primarily on the environmental conditions under which different mental constructs are assumed to operate and guide evaluation.

Parallel to these developments and largely unknown to their mechanistic counterparts, researchers from an intellectual tradition known as contextual behavioural science (CBS) have also sought to explain how people come to like and dislike stimuli. This approach draws on an alternative philosophical framework known as functional contextualism that makes no appeal to or *a priori* assumptions about mental constructs or their causal agency in behaviour and its change (Biglan & Hayes, 1996; Gifford & Hayes, 1999). Rather an exclusively functional epistemology is adopted, with behaviour defined as an interaction between the individual and environment that increases or decreases in probability as a function of its consequences. For several decades now contextual behavioural scientists have employed one functional account in particular - known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001) - in order to understand, predict and influence evaluative responding towards a variety of stimuli. In the second section of this chapter we shine a light on this

work and illustrate how in the absence of pairings, reinforcement or instructions, stimuli can come to acquire new or change their existing psychological functions via a behavioural process known as arbitrarily applicable relational responding. Finally, it is against this backdrop that the current work will be presented. Across a series of studies we highlight the implications and advantages of adopting a purely functional approach to the study of evaluation. We argue that such a perspective may equip researchers with novel procedures and theoretical frameworks as well as unlock previously unconsidered avenues of inquiry. In addition, we examine the possibility that although the mechanistic and functional frameworks operate at two independent levels of analysis, each may be mutually informed by the other, to the benefit of both¹.

1.1 A Mechanistic Approach to Evaluative Responding

Although evaluative responses towards stimuli in the environment can be established in a variety of ways, from mere exposure (Bornstein, 1989), to socialisation (Pettigrew & Tropp, 2006), descriptive information (Rydell & McConnell, 2006) and category membership (Pinter & Greenwald, 2004), cognitive and social psychologists have increasingly focused on evaluative conditioning as a means of manipulating this class of behaviour (see Gast, Gawronski & De Houwer, 2012 for an overview). Broadly speaking, evaluative conditioning (EC) refers to the finding that evaluative responses towards a previously neutral stimulus can be generated - or an existing stimulus altered - by pairing it with positive or negative stimuli. For example, contiguous presentations of an unknown Pokémon character with pleasant images results in that character being rated positively whereas pairing it with negative images results it being rated negatively (Olson & Fazio, 2001). Without doubt, the vast majority of work on EC has been conducted by researchers subscribing to a mechanistic world-view: this is evident in the methodological (De Houwer, 2011a), theoretical (Hofmann et al., 2010) and

¹ For purely stylistic reasons the pronoun “we” will often be used in place of “I” even though the work presented here is the product of a single doctoral candidate.

empirical focus of the field (De Houwer, Thomas & Baeyens, 2001). When this philosophical framework is adopted EC can be defined in three different ways; in terms of the *procedure* that gave rise to a change in liking, the *effect* generated by that procedure (i.e., the change in liking resulting from the pairing of stimuli), or the mediating mental *process* assumed to govern the change in liking resulting from the pairing of stimuli. In what follows we consider how mechanistic thinking has shaped our understanding of evaluative responding at each of these respective levels.

1.2 Evaluative Conditioning at the Procedural Level

When defined at the procedural level EC refers to arranging the environment in such a way as to pair stimuli, and as a result of these pairings, a change in liking takes place (De Houwer, 2011a). In a prototypical EC study, a neutral stimulus (CS) is repeatedly paired with a positive or negative unconditioned stimulus (US), and as a result, the evaluative functions of the CS are modified in-line with the US that it was paired with. Such procedures are argued to constitute a specific type of Pavlovian conditioning that focuses exclusively on changes in valence rather than any other change in response that occurs due to the pairing of stimuli. On the one hand, PC involves repeatedly pairing a biologically neutral stimulus (e.g., a tone) with a biologically relevant stimulus (e.g., food or electric shock) in order to establish a predictive relationship between environmental events (Rescorla & Wagner, 1972). Moreover, it typically involves presenting a CS before a US so that the CS actually predicts the subsequent presentation of the US. On the other hand, evaluative conditioning may involve pairing stimuli that have no biological significance and in a manner that does not necessarily involve a particular temporal sequence. For example, a fictitious brand product, such as chewing gum, may be paired sequentially or simultaneously with an image of a pleasant image (e.g., couple hugging) with a view to establishing that brand product as a positively valenced stimulus (Pleyers, Corneille, Luminet & Yzerbyt, 2007). Although

neither of the stimuli would be defined as biologically significant, the chewing gum is typically referred to as the CS and the pleasant image as the US (although strictly speaking the latter is more properly defined as a higher-order CS).²

Strong support for the claim that a change in liking resulted from the pairing of stimuli requires that the specific properties of the CS-US relation be identified. Researchers have therefore instantiated the core procedural property of EC in numerous ways with a range of stimuli and protocols. Overall, this work indicates 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), manner in which the CS-US relation was established (e.g., sensory preconditioning, higher-order conditioning; Hammerl & Grabitz, 1996; Walther, 2002), sources of current and historical contextual control (e.g., discriminative stimuli; Baeyens, Crombez, De Houwer, & Eelen, 1996; Gawronski, Rydell, Vervliet & De Houwer, 2010), subsequent 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 organism 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 involved directly relating the CS and US on the basis of spatio-temporal contiguity, changes in evaluative responding have also been obtained via other forms of relating, such as 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). Further complicating this picture is the fact that EC procedures sometimes present a successive stream of stimuli and require no overt response to the CS-US relation while at other times

² Although some EC studies do involve the presentation of biologically significant stimuli (e.g., Cacioppo, Marshall-Goodell, Tassinari & Petty, 1992), much work on EC does not make this distinction.

make progression through the task dependent on a stimulus-response contingency (Gast & Rothermund, 2011; Olson & Fazio, 2001; Walther et al., 2009).

When taken together, this work has provided valuable insight into the environmental conditions that serve to moderate changes in liking when stimuli are paired. These findings reveal that the CS-US relation is not a static or inflexible one but rather a dynamic relationship that it is situated in a wider environmental context. Changes in this context impact not only the magnitude, but direction and duration of evaluative responding. At the same time, the above research has also put to rest several early controversies surrounding EC, such as its apparent resistance to extinction (Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006) and insensitivity to contingency awareness (Walther & Nagengast, 2006). For instance, in a recent meta-analysis Hofmann and colleagues (2010) found that - similar to Pavlovian conditioning - the magnitude of EC effects are reduced when the CS is presented by itself following CS-US pairings (i.e., extinction) as well as significantly moderated by awareness of the CS-US contingency. Note, however, that the authors also acknowledge that unlike other forms of PC, many EC effects persist following the extinction of the CS-US contingency (Blechert, Michael, Williams, Purkis, & Wilhelm, 2008), and are influenced to a greater degree by the number of times that the CS and US have co-occurred than the statistical relation between those stimuli.

1.3 Evaluative Conditioning at the Effect Level

Although evaluative conditioning can refer to the manner in which stimuli are paired to produce a change in liking, it is also possible to examine this phenomenon at the level of the behavioural outcome or effect. Specifically, EC is the observed change in liking produced by the pairing of stimuli, regardless of what stimuli are paired, contextual factors manipulated or response registered. Although Pavlovian conditioning effects involve any potential

changes in responding that occur due to stimulus pairings, EC effects are argued to reflect a subclass of PC effects that only involve a change in one type of responding (i.e., liking).

When defined in this way, changes in liking that result from the pairing of (a) fictitious consumer products (Gast & De Houwer, 2012), cartoon characters (Olson & Fazio, 2001), nonsense words (Stahl & Unkelbach, 2009) and unknown individuals (Hütter, Sweldens, Stahl, Unkelbach & Klauer, 2012) with (b) valenced words (Walther, Langer, Weil & Komischke, 2011) or images (Corneille, Yzerbyt, Pleyers, Mussweiler, 2009) can be understood as EC effects. These changes in liking are not restricted to visual stimuli but have also been obtained when gustatory (Gast & De Houwer, 2012), olfactory (Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005), tactile (Hammerl & Grabitz, 2000) and auditory stimuli are used (van Reekum, van den Berg & Frijda, 1999). Although these effects have historically been calculated on the basis of direct measurement procedures such as categorisation performance or self-reported ratings, indirect tasks such as semantic and 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 so-called “automatic” or “implicit” evaluative responses to be captured and subjected to empirical scrutiny (see Nosek, Hawkins & Frazier, 2011). Although methodologically diverse, this latter class of procedures generally aim to provide an indirect estimate of automatic (evaluative) responding through (a) the speeded categorization of stimuli, (b) subjective judgments of ambiguous stimuli or (c) via physiological and neurological activity (for a detailed review see Gawronski & Payne, 2010).

Defining EC at the effect level also serves to delineate the similarities and differences between this class of behaviours and other types of behavioural effects. For example, changes in liking due to the successive presentation of a single stimulus (“mere exposure effect”;

Bornstein, 1989), pre-existing preferences for the related stimuli (Field & Davey, 1999), or even response-dependent contingencies (e.g., operant-conditioning; Beckers, De Houwer, & Eelen, 2002) are typically rejected as instances of EC. Rather, only those changes in responding that result from the pairing of stimuli can be considered members of this behavioural class. Overall, this work provides support for a genuine and heterogeneous set of evaluative responses that can be indexed in a variety of ways. The specific properties of any given effect appears to represent an interaction between (a) the individual's learning history with respect to the stimulus relation(s) (*see previous section*) and (b) the current and historical context in which those relations are embedded.

1.4 Evaluative Conditioning at the Mental Process Level

As noted above, the EC literature has almost exclusively been guided by researchers subscribing to a mechanistic world-view. Broadly speaking, mechanists conceptualise (psychological) events as being similar to a machine, composed of discrete parts that interact and are subject to specific operating conditions. The goal of psychological science from this perspective is therefore twofold (Bechtel, 2008). On the one hand, the basic mental operating processes that mediate between input (environment) and output (behaviour) must be identified. The constituent elements of a particular mental system can be described independently of one another and their fundamental structure remains the same when combined or in interaction with other mental constructs. They are often treated as ontologically "valid" such that the researcher's primary role involves developing an account of phenomena that actually exist and interact with behaviour. The truth or scientific value of a mechanistic model is therefore based on the *correspondence* between the mental construct it proposes and the set of behavioural observations that it aims to predict. In other words, cognitive researchers are primarily focused on the prediction of behaviour through the use of

theoretical models that bridge past events and current responses (see Hughes, Barnes-Holmes & Vahey, in press).

On the other hand, mechanistic researchers must also identify the operating conditions that are necessary and sufficient for these mental constructs to function. More precisely, mental processes are argued to operate under a restricted set of conditions that are separate from, but co-vary with, the environmental context under which behaviour is observed (Bargh, 1994; Moors, Spruyt & De Houwer, 2010). According to this perspective, different measurement procedures will provide a more or less sensitive index of different mental processes depending on whether their procedural properties correspond with the assumed operating conditions necessary for those processes to occur. For instance, researchers often treat the procedural properties of indirect procedures, such as speed or accuracy criteria, low task complexity and the absence of a requirement to self-report activity as either correlated with or equivalent to the assumed operating conditions necessary to observe automatic cognition and thus automatic behaviour (e.g., efficient, unaware, uncontrollable, unintentional). Likewise, the procedural properties that typically characterise direct procedures, such as an absence of speed or accuracy criteria, high task complexity or the requirement to verbally report activity, are also argued to correspond to the operating conditions necessary for controlled mental activity and thus controlled behaviour (e.g., non-efficient, aware, controllable and intentional; De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009).

When applied to evaluative conditioning, mechanism assumes that a single or set of mental constructs mediate between the pairing of stimuli (procedure) and the observed change in liking (behavioural outcome). Therefore the researcher's goal is to postulate mental theories 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 the

environmental conditions present within any given procedure. Broadly speaking, the majority of these mechanistic accounts can be sub-divided into two overarching categories³. The first, and currently dominant position is that the pairing of stimuli results in the automatic “bottom-up” formation of mental associations in memory and that these associations causally mediate changes in evaluative responding. For example, Martin and Levey’s (1994) holistic account suggests that the co-occurrence of the CS and US results in a mental association being formed between the two stimuli in memory. This association is a holistic one, comprised of properties of both stimuli in addition to the valence of the US. When the CS is subsequently encountered this holistic association may be activated along with the valence of the US it was previously paired with. Likewise, Jones and colleagues (2009) implicit misattribution account proposes that in situations where the CS and US co-occur, evaluative responses elicited by the US may be incorrectly attributed to the presence of the CS. In this instance, the pairing of stimuli will result in the formation of an associative, holistic representation comprised of the CS and valence of the US. Finally, the referential account (Baeyens, Eelen, Crombez & Van den Bergh, 1992) argues for two distinct forms of Pavlovian conditioning. While the first involves the formation of contingencies between the CS and US that signal a predictive relation between those stimuli, the second is referential in nature, with stimulus co-occurrences resulting in the formation of a mental link between the CS and US representations in memory. In the latter case, encountering the CS at a subsequent point in time serves to activate the US representation. However, because no predictive contingency

³ The distinction between EC and the PC at the mental process level varies dramatically from researcher to researcher. Several authors argue that PC and EC are driven by different mental mechanisms and thus are distinct types of learning (Baeyens et al., 1992; Jones et al., 2009) while others view EC as a specific type of PC that involves either stimulus categorisation (Davey, 1994), propositions (De Houwer, 2009a) or holistic representations (Martin & Levey, 1994). This picture is further complicated by the fact that mental models employ a large variety of terms when referring to similar phenomenon, such as “associative learning” (which can be interpreted as either respondent and/or operant learning), “classical conditioning”, “evaluative learning”, “Pavlovian conditioning”, “signal learning”, “referential learning”, “primitive learning”, “affective learning” and “conceptual learning”. In addition, the distinction between specific mental constructs such as associations and propositions often appears to supersede debates around the differences between PC and EC in the literature (see Gast et al., 2012).

was established between stimuli, specific features of the US (e.g., valence) are experienced without any expectation that the US itself will be encountered.

Although they differ in their predictions about what is being associated, as well as the conditions under which EC effects should emerge and change, these various models agree that once the CS and US are linked in memory, encountering the CS results in the automatic activation of properties of the US (for a review see Hofmann et al., 2010). In contrast, a second class of mental models have recently emerged that reject the associative position and propose that all forms of associative learning - including EC - arise due to the formation of propositions about the CS-US relation (De Houwer, 2009a; Mitchell, De Houwer & Lovibond, 2009). Whereas associations simply link the CS and US in some unspecified manner, propositions are qualified truth statements about organism-environment interactions that can differ in their precision and scope (e.g., “*the CS is opposite to the US*” or “*the CS is five times as positive as the US*”). According to this account participants utilise their knowledge of how stimuli are related as the basis on which to evaluate the CS (Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Gast & De Houwer, 2012). These propositions can be acquired in a number of ways, from prior knowledge and direct experience, to verbal instructions and deductive reasoning (De Houwer, 2009a).

In short, EC at the mental level of analysis involves identifying the mental processes and conditions that are assumed to mediate between the individual and environment. On the one hand, the referential, holistic, and implicit misattribution accounts share the assumption that contiguous presentations of the CS and US generate mental associations between these stimuli in memory. Although there are meaningful differences between these models, they each agree that subsequent presentations of a CS serves to activate the valence originally elicited by a US, which gives rise to a corresponding change in liking. These associations are thought to emerge quickly, in the absence of conscious awareness, attentional resources or

the intention to relate the CS and US (i.e., demonstrate many of the features of automaticity). On the other hand, propositional accounts substitute the notion of associations for truth qualified statements about the individual's interaction with the environment. Whereas associations simply convey the strength with which representations are linked in memory, propositions specify their strength, structure and content. These propositions are argued to require an awareness of the stimulus relation, as well as the time, cognitive resources and intention to relate those stimuli⁴.

1.5 Summary. Approaching the study of evaluation from an EC perspective has a number of important implications for our understanding of human likes and dislikes. First, methodological, theoretical and empirical attention is focused on respondent learning processes and preparations. For almost four decades now EC researchers have provided a rich and fruitful exploration of how the psychological functions of a stimulus can be altered by directly pairing it with other stimuli, either through experience, observation, instruction or inference. Although this emphasis on respondent preparations and processes is strategic (see De Houwer, 2007), it is important to appreciate that humans are not governed exclusively by respondent learning in their interactions with the environment. Rather other forms of learning such as stimulus generalisation, discrimination, and operant conditioning may also play a key role in how we come to like and dislike stimuli. Therefore a more sophisticated understanding of human evaluative responding would seem to require an account that can accommodate respondent processes as well as their interaction with increasingly complex forms of learning.

⁴ From a mechanistic perspective EC effects can - in principle - be mediated by any type or combination of mental constructs. 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 (Gawronski & Bodenhausen, 2011). This dual process account may explain why EC effects are sometimes obtained in the absence of contingency awareness or attention (i.e., associative processes) while at other times depend on those same factors (propositional processes). Nevertheless, such an account introduces additional complications such as how these two processes interact and singularly, interactively or additively produce EC effects.

Second, many of the EC effects outlined above appear to extend far beyond the scope of traditional respondent accounts. In particular, while EC is typically framed as a form of Pavlovian conditioning little or no reference is made to the important role that language seems to play in these evaluative responses. For example, EC preparations often involve the pairing of stimuli through verbal instructions, inferences, stories and statements, and even in studies where non-verbal stimuli are paired, participants are almost always exposed to a set of verbal instructions that indicate how they should respond on the task itself (e.g., “pay attention to the stimuli on the screen”, or “ignore certain stimuli on the screen”). If these EC (and thus PC) effects were strictly respondent in nature, as is often claimed, then non-verbal infants or non-human animals should also respond negatively towards novel individuals when informed that they “loath” cute kittens (Förderer & Unkelbach, 2012), or positively when those individuals are “friends” with a likeable person (Fiedler & Unkelbach, 2011). Yet this is clearly not the case. If, on the other hand, researchers argue that a history of language learning is needed to observe these types of EC effects, then they also need to specify a clear developmental explanation of how that language is acquired, what learning processes are involved and whether they influence or interact with respondent learning more generally. To date, the aforementioned (mental) models of EC have failed to articulate any such explanation.

Finally, it appears that mechanism offers little reward, rationale or means to study evaluative responding in a non-mental or mediational fashion. For instance, the learning processes and environmental regularities that influence the formation and change of EC effects are only of interest in so far as they specify when and how a given mental process or representation functions. In addition, the majority of theoretical accounts explain changes in evaluative responding (due to the pairing of stimuli) as the product of mental mediators such as holistic representations, implicit misattributions, conceptual categorisation, propositional

reasoning or referential learning. While this analytic strategy is fully in keeping with the assumptions, values and goals of mechanism, it is not without its own complications. In particular, mechanism requires that the researcher make *a priori* assumptions about how observable physical events relate to non-observable mental constructs and these assumptions are often fraught with complications (for a detailed review see Borsboom, Mellenbergh, & Van Heerden, 2004; De Houwer, 2011b; Hughes et al., in press).

In light of the above, and in the interest of a flexible and progressive science, it may be important to consider other approaches to the study of evaluative responding. Indeed, if complex human behaviour is always explored from a purely mechanistic position, then other useful and potentially productive theories, methodologies and findings may be missed. In what follows we outline an intensely pragmatic and non-dualistic alternative known as functional contextualism and illustrate how it has fostered a radically different understanding of evaluative responding than that offered thus far.

1.6 A Functional Approach to Psychological Science

While various forms of behaviourism have emerged over the last hundred years, the most empirically and theoretically productive contemporary branch is arguably that of contextual behavioural science (CBS; Hayes, Barnes-Holmes & Wilson, 2012). At the core of this intellectual tradition resides a philosophical framework known as functional contextualism which specifies (a) the assumptions, goals and values of the researcher, and by implication, (b) their observations, principles, theories and methodologies (see Hayes, 2004, Levin & Hayes, 2009). According to this perspective, (psychological) science involves a single unified goal; to understand, predict and influence behaviour with scope (explain a comprehensive range of behaviours across a variety of situations), precision (applying a restricted set of principles to any event) and depth (cohere across analytical levels and domains such as biology, psychology, and anthropology). Perhaps most importantly in the

context of the current chapter, functional contextualism differs in three dramatic ways from the mechanistic framework outlined above.

First and foremost, this approach adopts an exclusively functional epistemology. Instead of locating behavioural causation in the mind, scientific analysis is focused on the functional relations between the (past and present) environment and behaviour that unfold across both time and context. Consequently, any appeal to or *a priori* assumptions about hypothetical mental constructs or their causal agency in producing behaviour is omitted. There is no mechanistic requirement for spatially and temporally contiguous events to mediate the relationship between environment and behaviour. Rather, behaviour is defined as an on-going action that always occurs within and in response to a current and historical context. This context can “project outward spatially to include the entire universe...backward in time infinitely to include the remotest antecedent, or forward in time to include the most delayed consequence” (Hayes & Brownstein, 1986, p.178). Given that the temporal and spatial parameters of an environmental context can vary dramatically, functional researchers 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 over the behaviour of interest. When such an approach is adopted, evaluative responses are not explained in terms of mental constructs or mediating processes but as ongoing actions that either increase or decrease in probability as a function of their environmental determinants and consequences.

Second, the functional approach also differs from mechanism in how it conceptualises the environment. According to this perspective, the context that shapes and maintains behaviour can not only stretch backwards or forwards in space and time but also refers to the “internal environment” inside the person’s skin or the “external environment” outside the skin. Although early methodological behaviourists such as Watson (1924) focused

exclusively on public behaviour and excluded private events from legitimate analysis, contemporary functional researchers simply arrange behaviour along a continuum from public (e.g. walking, painting) to private (e.g. thinking, feeling, and remembering). It is worth noting here that private events are not “non-physical” but differ from their public counterparts only in their ease of accessibility (Skinner 1945; Hayes & Brownstein, 1986). Thus CBS views both public and private behaviours as dependent variables (i.e., effects for which we must find a cause) and environmental regularities external to the behaviour of interest as “independent variables” (i.e., the causes of behaviour).

Contextual behavioural scientists adopt this approach in order to achieve their unified goal of prediction-and-influence. To illustrate this more clearly, consider the associative models of evaluative conditioning outlined above wherein a change in liking due to the pairing of stimuli is mediated by the formation, activation or modification of mental links between associations in memory. Although these mental or non-physical events may be treated as the cause of a particular behaviour (e.g., ratings on a Likert scale) they are not open to direct manipulation. Instead they can only be inferred from relevant changes in behaviour produced through on-going interactions in and with the environment. On the one hand, if the researcher’s analytic goal involves only prediction, then emotions, thoughts, mental links, neurological activity or any other variable can be viewed as a “cause” of behaviour so long as it reliably precedes that behavior. On the other hand, if that same researcher wants to achieve both prediction-and-influence, appeals to any of the above explanations are ultimately unacceptable. In order to exert influence over behavior the researcher must successfully manipulate events external to that behavior, and only contextual variables located in the environment can be manipulated directly (see De Houwer, 2011b; Hayes & Brownstein, 1986). “Stated another way, analyses that deal only in dependent variables (e.g., emotion, thought, overt action) can never be fully adequate as measured against the pragmatic

purposes of functional contextualism” (Hayes, 2004, p.647). Consequently, scientific analysis is not complete until the causal variables external to the behaviour of interest have been identified – not because of some dogmatic adherence to monism but rather as a pragmatic means to achieve the goals of CBS⁵.

In short, although CBS accepts private events as a legitimate subject matter of scientific inquiry, it refuses to assign them causal status over behaviour. When scientific analysis allows for the use of mental constructs that cannot be directly manipulated, only prediction but not influence is attainable. Whereas prediction alone is acceptable according to the mechanistic framework, it is entirely unsatisfactory when measured against the goals of functional contextualism. Therefore, by recasting emotion, cognition and evaluation as a tendency to publicly or privately behave in a certain way within a certain context, functional researchers seek to understand these behaviours using the same principles that are used to understand their public counterparts. In doing so, empirical activity shifts away from the search for mediating constructs and orientates towards identifying the environmental variables that govern those behaviours.

Third and finally, contextual behavioural scientists start out by empirically identifying functional relations between behaviour and environment and then abstract these relations into overarching “principles” that are high in precision, scope and depth. Examples include 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 behaviours of varying complexity while maintaining coherence and parsimony” (Levin & Hayes, 2009, p.6). When researchers seek to explain a number of

⁵ This is not to say that predictive relationships between one behavior and another are not useful – they clearly are (see Hughes et al., 2012). Rather our point here is that behavior-behavior relations in the form of thought-action, emotion-action or action-action are always considered incomplete from a CBS perspective until they specify the environmental factors outside the behavior of interest that are available for direct manipulation.

behavioural events by weaving together a set of inter-related principles functional theories emerge. In contrast to the mental models noted above, these analytic-abstractive theories of behaviour make no claims regarding the existence of an independent reality nor do they seek to discover the ontological status of constructs such as associations or propositions. Instead a pragmatic and inductive position is adopted that qualifies a model or theory as “true”, successfully, meaningful or valid in so far as it “works” (i.e., is able to achieve both prediction and influence over behaviour, with precision, scope and depth; Hayes & Brownstein, 1986). This differs significantly from mechanism where a theoretical concept is “true” or “false” on the basis of public agreement about its correspondence with the behaviour of interest. In other words, when functional researchers develop a theory or model that divides the world into parts, it is in order to achieve some goal, rather than to reveal its underlying structure⁶.

To conclude, contextual behavioural scientists draw upon functional contextualism as a philosophical framework that shapes the observations they make, methodologies they develop and theories they construct. In contrast to mechanism, this approach substitutes the hypothetico-deductive postulation of mental mediators with empirically driven induction. Following the lead of Darwinian natural selection, CBS adopts a consequential conception of causality. Public and private behaviours are defined as actions that take place in a context and are selected or discarded according to their consequences for the organism. When viewed in this way, behaviour cannot be separated from the historical and current context in which it is situated. Rather, the researcher’s goal involves identifying functional relations between

⁶ A pragmatic truth criterion that emphasizes prediction and influence over behavior is adopted for another important reason. Given that the environmental context can in principle “project outward spatially to include all of the universe...backward in time infinitely to include the remotest antecedent, or forward in time to include the most delayed consequence” (Hayes & Brownstein, 1986, p.178), the notion of “successful working” provides functional researchers with an end-point to their analysis. Specifically, scientific activity need only proceed to a point where prediction-and-influence is - in principle - possible.

behaviour and environment, abstracting these relations out into overarching behavioural principles and subsequently weaving these principles together to form functional theories.

Over the past twenty years Relational Frame Theory (RFT; Hayes et al., 2001) has drawn upon a handful of interrelated behavioural principles in order to unite a wide range of verbal and cognitive phenomena under one theoretical umbrella. From self and perspective taking to intelligence, language and (implicit) cognition, functional researchers have shown that direct and derived stimulus relating play a key role many complex human behaviors. In what follows we outline the core features of this functional theory and examine how this approach may allow for a better understanding, prediction and influence of human evaluative responding than direct contingency (associative) accounts alone.

1.7 Relational Frame Theory

Relational Frame Theory proposes that the core defining property of human language and cognition is a type of behaviour known as arbitrarily applicable relational responding. At the core of this approach is the notion that relating is an action that involves responding to one event in terms of another, and that humans and non-humans alike can learn to relate stimuli in a number of different ways (e.g., respondent learning, operant learning, stimulus generalization and discrimination). In the case of respondent and operant learning, the organism learns to discriminate the relation between stimuli based on a directly trained contingency previously encountered in its learning history. For instance, mammals, birds, fish and insects can be trained to form relations between stimuli based on their physical properties (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001; Harmon, Strong & Pasnak, 1982; Reese, 1968) or even relate arbitrary stimuli such as abstract shapes and symbols based on their shared functions (Vaughan, 1988). In a similar manner, stimulus generalization and discrimination also require a prior history of relating but with the additional requirement that the previously encountered stimuli bear a physical similarity to the stimulus being

generalized or discriminated. Interestingly, over forty years of research indicates that non-human animals such as pigeons (Lionello-DeNolf, & Urcuioli, 2002), chimpanzees, and baboons (Dugdale & Lowe, 2000, Hayes, 1989; Sidman et al., 1982) are restricted to relating based on a direct history of training. In situations where nonhumans show novel relational performances these seem to be restricted to specific procedural parameters of the relational task, and/or to the formal properties of the stimuli involved (see Lionello-DeNolf, 2009).

Unlike many of their counterparts in the animal kingdom, humans develop the ability to spontaneously derive novel relations between objects and events that were never directly trained or instructed, and without regard to specific procedural parameters and the formal properties of the related stimuli. RFT defines this ability to derive novel relations between and among different stimuli in the absence of direct training as arbitrarily applicable relational responding. A sizable body of work now indicates that this type of behaviour is an overarching purely functional type of operant response class that is learned early on in our development through interactions with the verbal community and is defined according to the presence of three core properties: mutual entailment, combinatorial entailment, and the transformation of stimulus function (see Hayes et al., 2001; Rehfeldt & Barnes-Holmes, 2009)⁷.

To illustrate these three properties imagine that an individual is taught, through either direct contingency learning or instruction, that one stimulus (A) is the same as a second stimulus (B) and B is the same as C. In this scenario, mutual entailment refers to the bi-directional relation that emerges between two stimuli in the absence of explicit training. In other words, if A is the same as B, then humans will also derive a second relation (that B is the same as A) without any additional training. Combinatorial entailment refers to the

⁷ While humans are phylogenetically prepared to learn associations between stimuli (respondent learning) and to be governed by the consequences of their actions (operant learning) the ability to derive the relation between stimuli appears to be learned behavior that emerges through a history of generalized operant responding (for a detailed treatment of how derived relational responding emerges see Törneke, 2010).

functional relations that emerge between two or more mutually entailed stimuli. Thus, if A is bigger than B and B is bigger than C, then humans will spontaneously derive that A bigger than C as well as C smaller than A. Finally, once stimuli have been mutually or combinatorially related to one another, the (psychological) functions of those stimuli may be transformed in accordance with the stimulus relation. Imagine for instance that an aversive function is established for the A stimulus (e.g., a shock), and A is then related as equivalent to a number of other stimuli (e.g., B, C, D). Given appropriate contextual cues, these other stimuli will also acquire the negative functions of A despite the fact that they were never directly paired with a shock (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994). What is important to note here is that a transformation of function will always depend on the relation established between stimuli. Thus, if an opposition relation is established between A and B, and A is then paired with shock, the fear arousing functions of A will not necessarily transfer to B. Rather the emotional functions of B may come to be transformed in-line with the stimulus relation without direct training or instruction to do so. This may explain why humans can find stimuli directly paired with unpleasant events as pleasurable or reinforcing when an opposition relation is formed (Whelan & Barnes-Holmes, 2004).

We now know that the way in which stimuli are related, as well as psychological functions transformed through those relations, is determined by two forms of contextual control. On the one hand, stimuli can be related to one another in a vast number of ways, from simple mutually entailed relations between single stimuli to combinatorial relations involving multiple stimuli, to the relating of stimulus relations to other relations (often termed relational networks) to the complex relating of entire relational networks to other networks. In each of the above cases, the relation between stimuli may be based on equivalence (Cahill et al., 2007), similarity and opposition (Dymond, Roche, Forsyth, Whelan & Rhoden, 2008), hierarchy (Gil, Luciano, Ruiz & Valdivia-Salas, 2012), comparison (Vitale, Barnes-Holmes,

Barnes-Holmes & Campbell, 2008), temporality (O’Hora et al., 2008) and/or causality. Relational responding may also include deictic or perspective-taking relations (McHugh, Barnes-Holmes & Barnes-Holmes, 2007). On the other hand, a wide range of psychological functions can be transformed through these relations, including discriminative (Dougher, Hamilton, Fink & Harrington, 2007), affective (Barnes-Holmes, Barnes-Holmes, Smeets & Luciano, 2004), approach (Gannon, Roche, Kanter, Forsyth & Linehan, 2011), avoidance (Roche, Kanter, Brown, Dymond & Fogarty, 2008), sexual (Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000) and extinction functions (Dougher et al., 1994).

In short, Relational Frame Theory posits a rather simple notion – that complex human behaviour reflects the learned and contextually controlled ability to arbitrarily relate one stimulus to another. Throughout much of the past two decades this basic idea has taken root and flourished into a coherent, parsimonious and progressive account with a strong empirical foundation. Seemingly disparate and unrelated phenomena such as language (Hayes et al., 2001), intelligence (Cassidy, Roche, & Hayes, 2011) and implicit cognition (Hughes et al., in press), not to mention self and perspective taking (McHugh & Stewart, 2012) have all been predicted and influenced on the basis of the aforementioned theory. Moreover, much of this progress at the basic research level has fed into and produced positive influences in applied domains such as psychopathology (Gaudiano, 2011), developmental disability (Rehfeldt & Barnes-Holmes, 2009), education (Fox, 2006), and cultural change (Wilson, Hayes, Biglan & Embry, in press).

With respect to human evaluative responding, a burgeoning RFT literature indicates that when stimuli are related to one another in different ways, they may come to acquire new or change their existing (evaluative) functions without the need for direct stimulus pairings. More precisely, it appears that the ability to respond in an arbitrarily applicable fashion impacts upon every other known behavioural principle and forever changes how we interact

with the world (and by implication how we come to like and dislike stimuli). If correct, then derived stimulus relating has important ramifications for the evaluative conditioning research outlined above. As we shall see, contiguous presentations of stimuli in space and time may no longer simply involve respondent learning once a history of arbitrarily applicable relational responding is in place. Rather these stimuli may also be knitted together into derived relations without any training or instruction to do so.

1.8 Evaluative Responding as Relational Responding

1.9 Respondent Preparations

Similar to their EC counterparts, a number of RFT researchers have used respondent-like procedures to pair stimuli, and in doing so, alter their respective functions. Interestingly, much of this work has found that even when stimuli are simply paired on-screen, a number of novel and untrained derived stimulus relations tend to emerge (Leader & Barnes, 1996; Smeets, Leader, Barnes, 1997; Smyth, Barnes-Holmes & Forsyth, 2006). More often than not, these studies employ a respondent-like procedure in which a stimulus (A1) appears on-screen for one second and is then removed. Following a brief intra-pair delay a second arbitrary stimulus is also presented and then removed from the screen (B1). A three second inter-trial interval then occurs prior to the presentation of the next stimulus pair. This procedure continues across a series of iterative training trials until a number of different stimulus pairings are presented, such as $A1 \rightarrow B1$; $A2 \rightarrow B2$; $B1 \rightarrow C1$ and $B2 \rightarrow C2$.

Throughout the task participants are instructed to observe a successive stream of stimuli and at no point are they required to emit any overt response. Following training, participants often show evidence for mutual and combinatorial entailment consistent with the previously observed stimulus pairings. For example, having observed $A1 \rightarrow B1$ and $B1 \rightarrow C1$ pairings, participants may match A1 to C1 and C1 to A1 without explicit reinforcement or instructions to do so.

Perhaps more importantly in the context of the current chapter, these stimulus pairing procedures can also lead to a derived transformation of evaluative functions. For instance, Smyth et al., (2006) established a negative emotional function for one stimulus (A1) and no function for a second stimulus (A2). Thereafter, a respondent-like procedure similar to that outlined above was used to generate four separate stimulus pairings (A1→B1, A2→B2, B1→C1, B2→C2). Following this training the authors found that the evaluative function established for A1 also emerged for B1 and C1. In contrast, participants did not respond in an evaluative manner towards the A2, B2 or C2 stimuli (for related findings see Tonneau & Gonzalez, 2004). These studies indicate that contiguous presentations of stimuli in space and time may certainly involve respondent *procedures* but not necessarily respondent *processes* – at least where verbally trained humans are concerned. Rather directly pairing a series of stimuli with one another (e.g., A-B, B-C) may cause stimuli that were never directly paired to be spontaneously woven together into derived relations (e.g., A-C and C-A).

1.10 Non-Respondent Preparations

As noted above, stimulus pairing procedures are only one way in which EC researchers have sought to generate and modify evaluative responding. Stimuli have increasingly been related on the basis of verbal information (De Houwer, 2006), “relational qualifiers” (Förderer & Unkelbach, 2011; Walther, Langer, Weil & Komischke, 2011) and inferences (Gast & De Houwer, 2012), while a number of studies have also sought to establish EC effects using stimulus-response contingencies (Gast & Rothermund, 2011; Dack, Reed & McHugh, 2010). Although the words in these instructions, stories and inferences have been treated as CSs and USs such procedures involve far more than the pairing of stimuli or basic respondent processes. According to RFT, these learning scenarios do not simply result in the pairing but rather relating of stimuli in a multitude of different ways via contextual cues such as “goes with”, “loves”, “hates”, and “is opposite to”.

Importantly, these contextual cues need not always be words; they may also take the form of other discriminative properties of the context such as background colour (Gawronski et al., 2010), or verbal rules specified by the researcher (Zanon, De Houwer, & Gast, 2012).

In many respects, these recent developments within EC mirror the basic empirical and conceptual agenda pursued by RFT researchers over the last two decades. An extensive body of work now indicates that stimuli can acquire new or change their existing evaluative functions by participating in derived stimulus relations, such as those found in the above stories, instructions and inferences. Critically, these changes in evaluative responding occur in the absence of direct pairings, reinforcement or instructions. To illustrate, consider the work of Barnes-Holmes and colleagues (2000) who established two equivalence classes consisting of an emotive word, non-sense syllable and fictitious brand product (i.e., *Cancer-Vek-Brand X* and *Holiday-Zid-Brand Y*). Following training, participants were then presented with two different samples of cola labelled “Brand X” and “Brand Y” that were, unknown to them, identical in taste. The authors found that those participants who passed a test for equivalence rated Brand Y as more pleasant than Brand X while those who failed the equivalence test did not report any significant preference for either brand. This derived transformation of function through equivalence classes has now been replicated numerous times and used to establish preferences for soft drinks (Smeets & Barnes-Holmes, 2003), investigate gambling behaviours in young children (Dymond, Bateman & Dixon, 2010), alter the emotional functions of related stimuli (Barnes-Holmes et al., 2004; Cahill et al., 2007) and to provide a functional explanation for the IAT effect (O’Toole, Barnes-Holmes & Smyth, 2007). At the same time, stimuli that participate in non-equivalence relations may have their evaluative functions altered in increasingly complex ways. For instance, humans can come to fear harmless stimuli more than those that were directly paired with aversive events such as a shock, and even avoid those harmless stimuli altogether, on the basis of

comparative (Dougher et al., 2007), similarity or opposition relations (Dymond et al., 2008). Moreover, a stimulus that was initially avoided because it participated in a derived relation may subsequently be approached when the aversive functions of other stimuli in that network are extinguished (Roche et al., 2008). When taken together, these findings suggest that although respondent procedures may pair stimuli on the basis of spatio-temporal contiguity, the obtained outcome seems to reflect an alternative learning process (i.e., arbitrarily applicable relational responding) where verbally trained humans are concerned. Indeed, people may approach or avoid stimuli that were never directly paired with appetitive or aversive events simply because they participate in derived relations. Thus the behavioural process of arbitrarily applicable relational responding in conjunction with other (respondent) learning processes may open the door to a more complete understanding of human likes and dislikes than direct contingency accounts alone. In other words, RFT seems to account for a wide spectrum of evaluative responses - from those originating in the simple contiguous presentations of stimuli with one another all the way up to those established via stories, observation, inference and instruction - by drawing on a number of inter-related behavioural principles.

1.11 Summary. The learned ability to respond on the basis of derived relations and transform the psychological functions of stimuli fundamentally alters how humans interact with the world. Once a history of arbitrarily applicable relational responding is in place humans are no longer limited to direct contingency learning like so many of their non-human counterparts. Rather, stimulus functions may be rapidly established, changed and/or extinguished in a near infinite number of ways. If Relational Frame Theory is “correct” (at least in a functional contextual sense of the word), then people are constantly awash in a sea of relating and this ability to respond relationally dictates and pervades every element of their lives, from early infancy right through to their deathbed. For instance, people may not simply

consider their friends to be positively or negatively “valenced” but relate them to a rich network of other stimuli, such as “nice”, “honest”, “athletic”, and “funny”. If they are subsequently informed that a new individual is the opposite of a friend they may immediately relate a whole host of stimuli (such as “nasty”, “liar”, “boring” and “ugly”) to that individual in the absence of any direct experience or instruction to do so. Likewise, preferred brands, celebrities and sports stars may not only be positive or negative; they may be related comparatively, hierarchically, or via coordination to stimuli such as “young”, “sexy”, and “cool”, or “crazy”, “weird” and “fantastic”. This capacity to relate one network of relations to another and transform the psychological functions through those networks may explain why people can feel “trapped”, “pressured” or “stuck” in relationships, careers and ways of thinking; come to be governed by relational responses such as “I should value X” or “A good person would do Y” and even terminate their own lives based on statements such as “When I die my suffering will stop, it will be a relief and the world will be a better place” (for a detailed discussion of these and related topics see Hayes et al., 2001; Torneke, 2010).

Stated more precisely, if RFT and derived stimulus relating are indeed useful ways of thinking about the world that ultimately lead to better prediction-and-influence of behaviour, then two important implications for the study of evaluation become apparent. On the one hand, a purely respondent account that makes no reference to these relational abilities may provide an insufficient or incomplete understanding of how humans come to like and dislike stimuli in their environment. Indeed, an extensive and rapidly growing body of work now indicates that the transformation of function effects outlined above cannot be explained by or reduced to direct contingency processes – rather a new behavioural principle is needed (Hayes et al., 2001). In other words, while derived stimulus relating is a generalized operant response (and thus appeals to nothing new at the level of behavioural principle), the

transformation of functions seems to be a new kind of stimulus control that cannot be readily subsumed under traditional respondent or operant accounts.

To illustrate this point more clearly, imagine that a boy is bitten by a dog while walking down a familiar street and later learns for the first time that another word for dog is “hond”. When subsequently playing on a different street one of his friends turns to the boy and says “oh, look a hond” and the boy begins to cry. In explaining this crying behaviour we cannot say that “hond” is a CS because, by definition, it does not have the history of a CS. Rather, it is only “CS-like” because of a learned pattern of derived stimulus relating in which the psychological functions established for one stimulus (dog) are transformed through that network to another stimulus (hond) based on the presence of specific contextual cues (in this instance the word “is”). Consequently, while a respondent account may provide valuable insight into the modification of psychological functions, it is only one piece of a larger puzzle as far as complex human behaviour is concerned. “If the transformation of function is accepted, as a new principle, then virtually everything the basic science of behaviour analysis has established for humans, through the study of nonhuman behaviour, has to be re-examined and possibly reworked. This is indeed a daunting task, but if the basic premise of RFT is correct, then psychology will need to face this challenge head on” (Barnes-Holmes, Hayes & Barnes-Holmes, 2012, p. 41). Evaluative responding is no exception to this argument. In contrast to the EC models identified before, RFT argues that changes in liking (even those established via the pairing of stimuli) reflect a learned and contextually controlled ability to arbitrarily relate stimuli to one another in increasingly complex ways.

At the same time, if evaluative responding is inherently relational as suggested, then this assumption should remain true irrespective of how fast or slow those behaviours are emitted. Consistent with this notion, one functional account of so-called “implicit” or “automatic” cognition has recently been offered in the form of the Relational Elaboration and

Coherence (REC) model (Hughes et al., in press). According to this perspective, the behaviour captured by direct and indirect procedures is fundamentally relational in nature and reflects (a) an interaction between the individual's learning history with respect to the targeted relations and (b) the specific features of the context in which they are assessed. Importantly, the REC model makes a number of clear and testable predictions with respect to "automatic" and "controlled" evaluative responding. For instance, it should be possible to obtain evidence for derived relational responding across a wide range of measurement procedures, from the relatively slow and carefully considered self-report task to speeded and time-based alternatives such as the IAT, affective priming and IRAP. Both the IAT and the IRAP measure participant response latencies toward pairings of stimuli and positive/negative attributions to determine "implicit" or "automatic" evaluations. Evaluation is interpreted on the basis that faster responding to particular stimulus relations or their converse indicates consistency with participants' own history of learning. For example, faster responding during trials that require participants to relate "Black People" with positive attributes compared to trials that require participants to relate "Black People" with negative attributes is taken to indicate a positive evaluative bias towards this racial group. Affective priming also involves time-based measures of participant responding. In the current research context, it was expected that experimentally established relations should vary in their complexity and the psychological functions transformed through those relations should be dictated by contextual cues present in the environment. In particular, and given appropriate contextual control, stimuli that participate in relations with appetitive or aversive stimuli should also be evaluated positively or negatively, regardless of the fact that they were never directly paired with those stimuli to begin with. In what follows we unpack a number of these assumptions and provide an experimental analysis of the learning histories and current contextual variables critical for generating evaluative responses.

1.12 Overview of Current Research

Although the transformation of functions appears to be an empirically robust phenomenon that holds across a variety of populations, stimulus domains, and methodologies, several empirical issues still require attention at this time. First, while a number of studies have altered the functions of stimuli through opposition and comparative relations (Dymond et al., 2008; Whelan, Barnes-Holmes & Dymond, 2006), the vast majority of work involving evaluative functions has focused on a single type of relating (i.e., equivalence; see Barnes-Holmes, Keane, Barnes-Holmes & Smeets, 2000; Smeets & Barnes-Holmes, 2003; Smyth et al., 2006). Second, many RFT studies have relied heavily on response rate (Dougher et al., 2007), skin conductance (Auguston, Dougher & Markham, 2000) and accuracy (Gil et al., 2012) in order to assess transformations of function effects within the laboratory. Although researchers have increasingly employed self-report questionnaires in this area (Barnes-Holmes et al., 2000; Barnes-Holmes et al., 2004; Smyth et al., 2006), the use of indirect procedures such as the IAT, IRAP, priming and event related potentials is still very much in its infancy (although see Barnes-Holmes et al., 2005; O’Toole et al., 2007). Indeed, in every transformation of function study where an indirect procedure has been used, only equivalence relations were trained and tested. Finally and despite the fact that stimulus relations can be established in multiple different ways, many of the aforementioned studies have adopted the Matching-to-Sample protocol as the sole means to train and test for arbitrarily applicable relational responding.

In light of the above, the current thesis represents an attempt to fill each of these respective gaps in the literature. Across six studies, a number of coordination, opposition and comparative relations were established using a conditional discrimination task. In Experiments 1-4, participants were first exposed to a newly developed relational training procedure in order to establish the functions of ‘Same’ and ‘Opposite’ for two contextual

cues. Thereafter, a series of trials were presented containing either two Pokémon characters (Chapters 2-3), the names of fictitious brand products (Chapter 4) or potential prizes (Chapter 5) in addition to the previously trained contextual cues. By providing differential reinforcement for the selection of a cue in the presence of a specific stimulus combination, a number of increasingly complex relations were generated. In order to index the transformation of evaluative functions through these relations, a number of direct and indirect procedures were used (e.g., IAT, IRAP and affective priming tasks). In each case and consistent with the predictions outlined above, “automatic” and “controlled” evaluative responding was found to be inherently relational in nature. Irrespective of the stimuli or methodology used, the obtained effects were not governed by the mere presentation of stimulus pairings, but rather the relation established between those stimuli by contextual cues. Based on the success of these findings, Experiments 5-6 sought to demonstrate a derived transformation of function through comparative (‘More than’ and ‘Less than’) relations and capture these effects using a combination of direct and indirect procedures. Once again, evidence for the relational nature of human evaluative responding was obtained. Following this work we present the final chapter in this thesis which considers the implications of the above findings for RFT and the REC model and outlines potential challenges and opportunities that lie ahead for researchers interested in the study of evaluation.

Before proceeding, one important point should be noted. In walking the tightrope between the functional and mechanistic approaches the current thesis will draw upon analytic strategies, methodologies and stimuli from both traditions. Indeed, the work presented here represents an active attempt to bridge the gap between these two literatures, and in particular, introduce the merits of a functional level of analysis to our mechanistic colleagues. Importantly, this endeavour is not without its own costs. On the one hand, traditional behaviour-analytic readers may lament the use of group-level statistics and “cover stories”,

the absence of single-subject designs or the type of stimuli and procedures used. On the other hand, their mechanistic counterparts may question the lack of mental theory testing or the esoteric procedures used to train and test derived stimulus relating. Although we cross intellectual boundaries at the methodological level, borrowing tools and preparations from both traditions, we always maintain a clear and separate distinction between these positions at the philosophical and theoretical levels of analysis. In doing so, the current work will attempt to provide the first test of the recently proposed functional-cognitive framework advanced by De Houwer (2011b). At its core, this account argues that the functional and mechanistic approaches may indeed operate at fundamentally distinct levels of analysis, but both can be mutually supportive, insofar as theoretical, methodological and empirical developments at one level can lead to advances at the other level. In adapting functional methods and theoretical perspectives to the study of human likes and dislikes, the current thesis aims to lay the groundwork for future productive exchanges between these two intellectual traditions.

Chapter 2: A Transformation of Functions through Mutually Entailed Relations as Measured by the IAT and Self-Report Tasks

As outlined in Chapter 1, the majority of evaluative conditioning research has sought to alter the functions of a stimulus by presenting it in close spatial and temporal proximity to another valenced stimulus. While many researchers continue to adopt this strategy, others have recently begun to pair the to-be-related stimuli in the presence of contextual cues that specify how those stimuli should be related. For instance, Förderer and Unkelbach (2012) found that pairing a CS and positively valenced US images in the presence of the contextual cue “loves” resulted in a standard EC effect. However, when those same stimuli were related in the presence of a different contextual cue (“loathes”) the effect was completely reversed (see also Fiedler & Unkelbach, 2011; Peters & Gawronski, 2011; Walther et al., 2011). In a similar vein, relating one stimulus (the non-word ‘enanwal’) with a second stimulus (the word “positive picture”) in the presence of the contextual cue “will be followed by” resulted in the non-word being evaluated positively on both direct and indirect procedures (De Houwer, 2006). Indeed, EC effects can even be found if participants are provided with verbal information that allows them to infer that a neutral and valenced stimulus were related during training (Gast & De Houwer, 2012; see also Lovibond, 2003).

Within the EC literature, the above effects continue to be defined in terms of Pavlovian or respondent learning despite the fact that they are contingent on language and involve relating verbal stimuli. According to Relational Frame Theory, however, the above evaluative responses do not reflect simple respondent processes but instead the learned and contextually controlled ability to arbitrarily relate one event to another. From this perspective the participants’ history of arbitrarily applicable relational responding that they bring with them into the experiment influences the manner in which they relate stimuli. For example, in a typical EC preparation where two stimuli repeatedly co-occur, verbally trained participants

may treat ‘spatio-temporal contiguity’ as a basic contextual cue for relating a CS and US (e.g., “*CS goes with US*”, “*CS is related to US*”, “*CS is the same as US*”; see Smyth et al., 2006). However, when alternative forms of contextual control are introduced responding may come to be governed by such cues. Examples include pre-existing words such as “*more than*”, “*predicts*” and “*is opposite to*” or even background colours, non-sense syllables or shapes for which similar relational functions have been established in the individual’s learning history (Gawronski et al., 2010; Zanon et al., 2012). As such, we argue that it may be more accurate to conceptualize changes in human evaluative responding (even those resulting from the pairing of stimuli) as a transformation of function though derived relations that is under a specific form of contextual control, rather than an instance of basic respondent learning.

2.1 Experiment 1

With the above in mind, the current study sought to demonstrate the history of learning by which contextual cues (like the words used in Förderer & Unkelbach, 2011a or Walther et al., 2011) come to acquire their psychological functions, and as a result, modify the functions of stimuli that they are related to. More precisely, Experiment 1 set out to show that the functions of a stimulus may depend more on the derived relation it participates in than simply the stimulus it was previously paired with. Towards this end, two arbitrary shapes were randomly selected and the relational function ‘Same’ was established for one and ‘Opposite’ for the other. While previous work has often employed cues that were well-established in the verbal history of the individual (De Houwer, 2006; Peters & Gawronski, 2011) we generated those cues within the experiment itself using a modified “operant” version of the Picture-Picture paradigm (Levey & Martin, 1975). Thereafter, we established four separate stimulus relations by requiring participants to select one of the two cues when simultaneously presented with a Pokémon character (CS) and valenced image (US) (e.g.,

Pokémon1-Same-Positive; Pokemon2-Opposite-Positive; Pokemon3-Same-Negative and Pokemon4-Opposite-Negative). Following recent work by Stahl and Unkelbach (2009), as well as multiple exemplar training typically seen in RFT research, each of the four Pokémon characters were presented with multiple USs (rather than a single US) of the same valence. This strategy was adopted in order to ensure that participants abstracted the US images into two overarching classes of “positive” and “negative” stimuli that were related to a specific CS in the presence of a contextual cue.

If this training is successful, then the degree to which a Pokémon character is rated as positive or negative should be determined by the cue that governed the stimulus relation during training. For example, if a coordination (‘Same’) relation is established between a Pokémon and positive images a transformation of function from one stimulus to another should be evident (i.e., the Pokémon should be evaluated positively). However, if an opposition relation is established between those same stimuli then that Pokémon should be evaluated negatively - regardless of the fact that it was exclusively and repeatedly paired with positive images. Put another way, it should be possible to bring evaluative responding under relational rather than strictly associative control.

The current study also provided an opportunity to test a core assumption of the Relational Elaboration and Coherence (REC) model - namely - that direct and indirect task outcomes reflect the same behavioural process (derived relational responding) operating under different environmental conditions. In very general terms, this functional account suggests that responses typically defined as “automatic” or “implicit” frequently involve brief and immediate relational responses (BIRRs) while their “explicit” or “controlled” counterparts involve extended and elaborated relational responding (EERRs) (for a more detailed treatment see Hughes et al., in press). With respect to the current work, the REC model predicts that generating a series of derived relations between Pokémon characters and

valenced images should give rise to contextually controlled patterns of evaluative responding on both direct and indirect tasks alike. In order to test this assumption, participants not only provided evaluative ratings of the four Pokémon characters but also completed two separate Implicit Association Tests (IAT; Greenwald et al., 1998) – one targeting relational responses towards *Pokémon1-Same-Positive* relative to *Pokémon3-Same-Negative* and a second IAT assessing *Pokémon2-Opposite-Positive* relative to *Pokémon4-Opposite-Negative*. If performance on the IAT is driven primarily by simple Pavlovian or respondent learning, then Pokémon 1 and 2 (both paired with positive images) should be evaluated as positive while Pokémon 3 and 4 (both paired with negative images) should be evaluated as negative. However, if responding is under relational rather than associative control as predicted, then self-reported and IAT performances should be governed by the contextual cue with which the CS and US were related during training. In other words, participants should respond positively towards Pokémon 1 and 4 and negatively towards Pokémon 2 and 3 on both direct and indirect tasks. To our knowledge, no published study has ascertained whether mutually entailed coordination and opposition relations lead to differential outcomes on direct and indirect procedures.

2.2 Method

Participants and Design

Ninety one undergraduates (57 female), ranging in age from 18 to 33 years ($M = 20.8$, $SD = 4.1$) completed the study in exchange for a chocolate bar. All participants reported normal or corrected to normal vision. Overall, a 4(*Stimulus relation*: same-positive, same-negative, opposite-positive and opposite-negative) x 2(*Training*: one versus two sessions) design was used, with the latter variable manipulated between-participants. Two additional method factors were also counterbalanced: IAT presentation order and IAT block order. Data

from eleven individuals who failed to meet the training criteria were excluded from subsequent analyses (*see below*).

Materials

Stimuli. Prior to the study, thirty one (non-participating) undergraduates were presented with a set of twelve CS images (Pokémon) and asked to provide an evaluative rating for each using a scale ranging from -3 (Negative Feelings) to +3 (Positive Feelings) with 0 as a neutral point. From these twelve images the four Pokémon characters deemed the most neutral were selected (*piloswine, drifblim, baltoy and lileep*; *Mean rating = .054, .081, .162 and .216* respectively). The USs consisted of 18 pleasant and 18 unpleasant pictures selected from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2005). All images were approximately 9.5 by 9.5cm in size and displayed on either the upper right or left side of the computer screen. Mean evaluative ratings were 2.1 for the set of positive pictures and -2.5 for the set of negative pictures, $t(28) = 22.7, p = .001$. Two arbitrary symbols (i.e., \mathcal{Q} and \mathcal{A}) were used as contextual cues while a computer program written in VB.Net controlled the presentation of stimuli and recorded responses.

IATs. Brief and immediate relational responding was indexed using two IATs. In either case, images of two Pokémon characters with their names printed underneath served as one set of target stimuli and the words “Good” and “Bad” as another. Eight positively valenced and eight negatively valenced adjectives served as one set of attribute stimuli (positive: *happy, love, pleasure, cheer, joy, kind, friendly, wonderful*; negative: *sick, horrible, agony, grief, terrible, awful, pain, sadness*) while images of the Pokémon characters at different orientations served as another set.

Ethics

The research procedures outlined below did not involve risk to participants and ethical considerations were focused primarily on informed consent, confidentiality and data

protection. Experiments 1-6 were relatively brief - taking less than one hour to complete - and therefore did not greatly inconvenience participants. At some points chocolate was provided to participants as a token of appreciation for their time (Experiments 1-3) while a small sum of money was provided in Experiments 5-6 for experimental reasons (see Chapter 5 for further details). Additionally, it should be noted that in Experiment 4 participation was offered as a course work option to undergraduate students. However, an alternative option was provided so that students who did not wish to participate were free to complete an essay instead. Students were informed they were under no obligation to participate and there would be no penalty of any description should they choose not to do so.

Procedure

Upon arrival to the laboratory participants were welcomed by an experimenter. After they had given their informed consent (see Appendix A), they were seated in front of a computer screen on which all instructions were presented. For each participant the study consisted of the following three phases: contextual cue training, relational training, followed by direct and indirect measures of evaluation.

Contextual Cue Training

We first sought to establish the relational functions of ‘Same’ and ‘Opposite’ for two arbitrary nonsense symbols using a modified “operant” version of the Picture-Picture paradigm. On-screen instructions informed participants that the computer would present a series of trials containing two pictures at the top of the screen and two symbols at the bottom of the screen. Selecting the symbol on the lower left side of the screen by pressing the ‘f’ key indicated that *“you are saying that this symbol describes the relationship between the two pictures at the top of the screen”*. Alternatively, *“If you choose the symbol on the lower right side of the screen using the ‘j’ key you are saying that this symbol describes the relationship between the two pictures at the top of the screen.”* Participants were asked to take their time

throughout the task and try to respond as accurately as possible (see Appendix B). Thereafter the researcher left the room and training began.

Contextual cue training consisted of two different types of trials – those designed to establish the relational function of similarity for one symbol and those designed to establish the relational function of opposition for the other (the relational functions assigned to the two symbols were randomly counterbalanced across participants). On each trial, two pictures bearing a non-arbitrary relation of either similarity (e.g., two images of dogs) or opposition (e.g., an image of winter and summer) were presented at the top of the screen and the two symbols were positioned on the bottom of the screen (see Figure 2.1). On ‘similarity’ trials choosing the symbol to be trained as ‘Same’ in the presence of two pictures bearing a non-arbitrary relation of similarity was reinforced. For example, when presented with two images of fire, selecting the ‘Same’ cue caused the written feedback “Correct” to appear and remain in the middle of the screen for 1000ms. All stimuli then disappeared and following a 1000ms inter-trial interval the next trial began. In contrast, selecting the incorrect symbol on similarity trials - such as the ‘Same’ cue when presented with a picture of fire and water – caused the written feedback “Incorrect” to appear in red. This feedback remained on screen until the participant emitted the correct response. Thereafter, all stimuli disappeared, followed by the inter-trial interval and subsequent trial. A broadly similar pattern of responding was required for ‘opposition’ trials. Critically however, selecting the ‘Opposite’ cue when presented with pictures bearing a non-arbitrary relation of opposition was reinforced (e.g., fire/water, black/white square, rich/poor person) while selecting the ‘Same’ cue was punished. In each case, progression to the next trial was contingent on selecting the correct symbol.

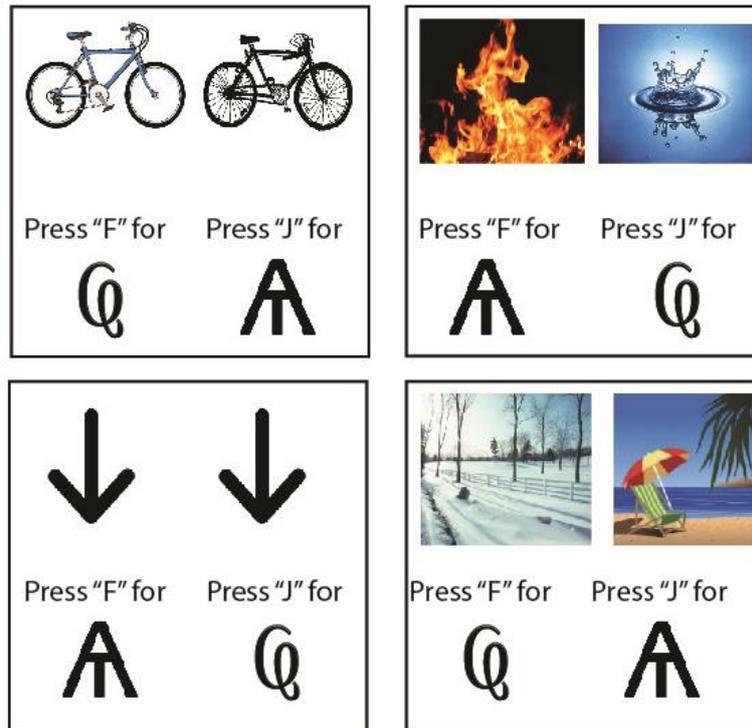


Figure 2.1. Examples of the similarity and opposition training and testing trials. Each trial consisted of two pictures at the top of the screen and the two to-be-trained contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).

Participants were exposed to a minimum of one and a maximum of three blocks of 50 training trials. Within each block the allocation of the two symbols to the lower left and right sides of the screen, as well as presentation of similarity and opposition trials was varied in a quasi-random order, with the constraint that a symbol could not occupy the same location, or a specific trial-type be presented, on more than three consecutive occasions. To successfully complete the training phase a mastery criterion of 100% accuracy across 20 successive trials was required. Thereafter, participants were exposed to a block of test trials to ensure that the cues acquired the relational functions of ‘Same’ and ‘Opposite’ as intended, and not simply through compound stimulus control (see Dougher, Perkins, Greenway, Koons & Chiasson, 2002). These test trials were identical to those encountered during training with two notable exceptions. First, an entirely novel set of images bearing either a relation of similarity or opposition were presented, and once again, participants were required to select one of the two

contextual cues. Second, no feedback was provided when either a correct or incorrect response was emitted. In order to successfully complete the test block and progress to the next stage of the study, the correct contextual cue had to be selected on at least 20 out of 24 trials. While this criterion is an arbitrary cut-off, it is highly unlikely that participants could have achieved this level of performance without learning the contextually-controlled relational responses. Failure to do so resulted in re-exposure to another set of training and testing blocks until either (a) the mastery criterion was met or (b) three training and testing blocks were completed. Attaining the mastery criteria resulted in progression to relational training while failure to do so resulted in participants being thanked, debriefed and their data discarded (8 individuals out of a total of 91 were removed on this basis).

Relational Training

Using the previously trained contextual cues, four different stimulus relations between Pokémon characters (CS) and valenced images (US) were established. On-screen instructions informed participants that during the next section of the study, their goal was to determine the relationship between the two pictures at the top of the screen using the two symbols they had previously encountered in the experiment. Participants were asked to take their time and try to respond as accurately as possible (see Appendix C).

As before, training consisted of a minimum of one and a maximum of three blocks of 50 trials. Within a block, each trial displayed a single Pokémon character and positive or negative image at the top of the screen and the two contextual cues at the bottom. By differentially reinforcing the selection of one of the two cues in the presence of a certain stimulus combination, two coordination ('Same') and opposition relations were formed (i.e., *Pokémon1-Same-Positive*; *Pokémon2-Opposite-Positive*; *Pokémon3-Same-Negative*; *Pokémon4-Opposite-Negative*). For instance, when Pokémon 1 and a positive image, or Pokémon 3 and a negative image were displayed, selecting the 'Same' cue caused "Correct"

to appear and remain on screen for 1000ms. Thereafter, all the stimuli disappeared, and following a 1000ms inter-trial interval, the next trial commenced. In the case of an incorrect response, such as choosing ‘Opposite’ when presented with the above stimuli, an error message (“Incorrect”) was displayed and progression to the next trial was contingent on selecting the correct symbol. In contrast, when Pokémon 2 and positive images, or Pokémon 4 and negative images were presented, selecting the ‘Opposite’ cue caused “Correct” to appear on-screen while selecting the ‘Same’ cue caused “Incorrect” to appear (see Figure 2.2).

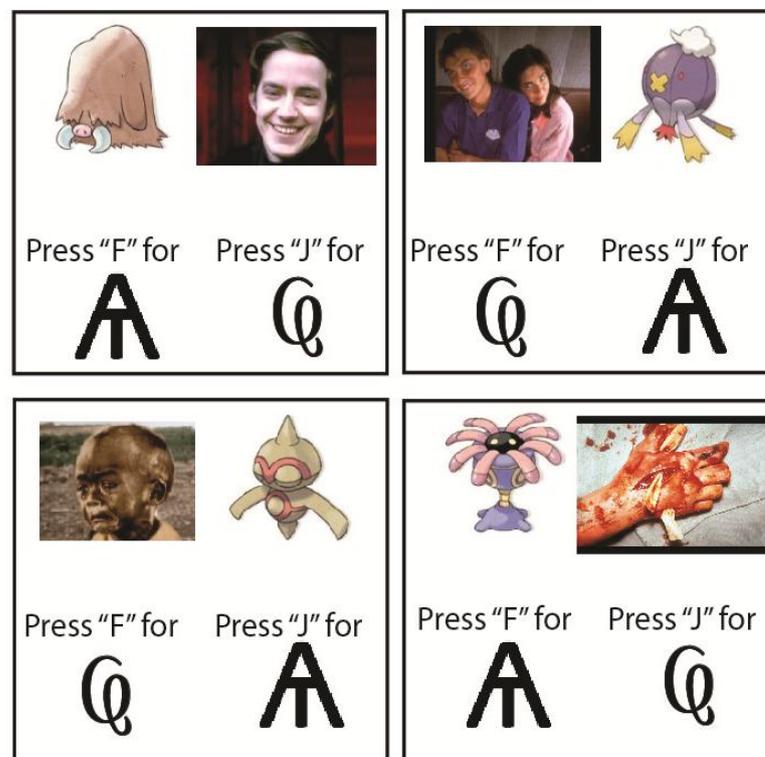


Figure 2.2. Examples of the four relational training and testing trials involved in establishing either a coordination or opposition relation between a Pokémon (CS) and valenced image (US). Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).

As before, participants were required to achieve 100% accuracy across 20 successive trials in order to progress to the test phase. Testing involved presenting the same Pokémon characters with novel, previously untrained images of the same valence. Once again, participants were required to select the correct contextual cue on at least 20 out of 24 trials.

Failure to achieve these criteria following three iterations of training and testing resulted in the participant being thanked, debriefed and their data discarded (three participants were removed on this basis). Throughout the relational training task each of the four Pokémon was presented with seven different images of the same valence (e.g., Pokémon 1 and 2 were always presented with positive images but never negative images and vice versa for Pokémon 3 and 4). In addition, the assignment of the four Pokémon to a contextual cue-US combination was counterbalanced across participants. Finally, to determine whether the amount of training received influenced evaluative responding, half of the participants received a single session of contextual cue and relational training while the other half were exposed to two consecutive sessions of training.

Indirect Procedure

In order to assess brief and immediate relational responding towards the coordination relations independently from the opposition relations, each participant completed two separate IATs. On the ‘coordination’ IAT evaluative responding for *Pokémon 1-Same-Positive* relative to *Pokémon 3-Same-Negative* was assessed while the ‘opposition’ IAT indexed responding towards *Pokémon 2-Opposite-Positive* relative to *Pokémon 4-Opposite-Negative*.

Prior to the onset of their first IAT, participants were informed that a series of images and words would appear one-by-one in the middle of the screen and that their task was to categorize those stimuli with their respective target (Pokémon characters) or attribute categories (“Good” and “Bad”) as quickly and accurately as possible. They were also informed that the target and attribute labels would appear on the upper left and right sides of the screen and correspond to either the left key (‘E’) or right keys (‘I’) respectively. Each trial started with the presentation of a fixation cross for 200ms in the middle of the screen followed immediately by an attribute or target stimulus. If the participant categorized the

image or word correctly - by selecting the appropriate key for that block of trials - the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red 'X' which remained on screen until the correct key was pressed. Overall, each participant completed seven blocks of trials. On the coordination IAT, the first block of 20 practice trials required the categorization of Pokémon characters with their super-ordinate category labels, with Pokémon 1 assigned to the left ('E') key and Pokémon 3 with the right ('I') key. On the second block of 20 practice trials participants categorized positive attribute stimuli with the category label "Good" using the left key and negative attribute stimuli with "Bad" using the right key. Blocks three and four (60 trials) involved a combined assignment of target and attribute stimuli with their respective category labels. Specifically, participants categorized Pokémon 1 and 'positive' words using the left key and Pokémon 3 and 'negative' words using the right key. The fifth block of 20 trials reversed the key assignments, with Pokémon 1 now assigned to the right key and Pokémon 3 with the left key. Finally, the sixth and seventh blocks (60 trials) required participants to categorize Pokémon 1 with 'negative' words and Pokémon 3 with 'positive' words. The opposition IAT was identical in all respects, with the exception that Pokémon 2 and 4 served as the target and attribute stimuli.

Before continuing it is worth noting that within each IAT the critical order of test block presentation was counterbalanced such that half of the participants completed the procedure in a "consistent-first" fashion while the remaining half completed it in an "inconsistent-first" fashion. For consistent-first participants, blocks three and four involved categorizing the Pokémon in a similar manner to prior training (e.g., Pokémon 1 with positive and Pokémon 3 with negative stimuli) while inconsistent-first participants were required to relate those same stimuli in opposition to what they previously learned during training (e.g., Pokémon 1 with negative and Pokémon 3 with positive stimuli). If coordination and

opposition relations were established in-line with prior training then response latencies should be shorter during consistent relative to inconsistent IAT trials. Specifically, faster responding should be obtained when participants have to relate Pokémon 1 or 4 with positive relative to negative stimuli and Pokémon 2 or 3 with negative relative positive stimuli.

Direct Procedures

CS ratings. Self-reported evaluative responding was assessed separately for each of the four Pokémon using a series of Likert scales. On each trial, participants were presented with one of the characters and asked to indicate their general impression of that stimulus using a scale that ranged from -9 (Negative Feelings) to +9 (Positive Feelings) with 0 as a neutral reference point (see Appendix D).

Contextual cue meaning. To ensure that the relational functions of the contextual cues were established in-line with experimental expectation, the two symbols were presented simultaneously on-screen and participants asked to indicate their respective meanings (see Appendix E).

Demand compliance. Finally, participants were asked whether they had intentionally rated the Pokémon in-line with what they believed the experimenter wanted or according to what they had previously learned about the characters (see Appendix F).

2.3 Results

Data Preparation

Contextual cue meaning. Seventy five participants (94%) reported the relational functions of the two contextual cues in-line with experimental expectations. On the one hand, forty seven participants (63%) rated the ‘Same’ cue as meaning “same”; eighteen participants (24%) rated it as meaning “similar” while the remaining 10 participants (13%) used a variety of terms “alike”, “linked” or “related”. On the other hand, thirty nine participants (52%) rated the ‘Opposite’ cue as “opposite”; twenty six (35%) rated it as meaning “different” while the

remaining 10 participants (13%) used a variety of terms such as “dissimilar”, “not related” or “not the same”. The data for the four participants who reported incorrect relational functions for the two cues were removed prior to analysis. That said, reanalyzing the data with these participants included did not change any of the statistical conclusions reported below.

Demand compliance. Two of the eighty participants who completed the study reported that they intentionally responded to the Pokémon in-line with the presumed expectations of the experimenter. The data for both participants were removed (although once again, reanalyzing the data with these participants included did not influence the obtained effects).

Preliminary analyses. Counterbalancing the relational functions ‘Same’ and ‘Opposite’ across the two symbols, Pokémon across the US-contextual cue combinations, the amount of training received and the order of the two IATs produced no significant effects. Consequently, analyses were collapsed across these various factors.

Self-Reported Ratings

To investigate the predicted changes in evaluative responding we calculated four mean likeability ratings, one for each of the Pokémon as a function of the contextual cue and valenced images that it was paired with (see Figure 2.3). Visual inspection of the graph revealed that ratings varied according to the type of stimulus relation established. For instance, when a Pokémon was related to positive images using the ‘Same’ cue (or negative images using the ‘Opposite’ cue) it was evaluated positively. However, this effect was completely reversed when Pokémon were related to negative images using the ‘Same’ cue (or positive images using the ‘Opposite’ cue). Self-reported ratings also appeared to increase in magnitude when they were generated via coordination rather than opposition relations, such that *Pokémon-Same-Positive* received larger positivity scores than *Pokémon-Opposite-*

Negative while *Pokémon-Same-Negative* was rated more negatively than *Pokémon-Opposite-Positive*.

This description of the data was supported by the results of a one way repeated measures analysis of variance (ANOVA). As predicted, a significant effect emerged for likeability score, $F(3, 74), = 101.1, p = .001, \eta^2_{\text{partial}} = .58$, with ratings of a Pokémon character varying according to the valenced images and contextual cue it was paired with. Post-hoc tests confirmed that when coordination relations were involved, Pokémon acquired the same functions as the valenced images they were paired with. Specifically, a significant difference emerged between *Pokémon-Same-Positive* and *Pokémon-Same-Negative* ($p = .001$), such that selecting the ‘Same’ cue when presented with a Pokémon and positive images resulted in that stimulus being evaluated positively ($M = 6.5, SE = .4$) whereas selecting the ‘Same’ cue when presented with a Pokémon and negative images resulted in that stimulus being evaluated negatively ($M = -5.8, SE = .5$). Critically, the direction of evaluative responding was completely reversed when opposition relations were established between those stimuli. In this case, a significant difference emerged between *Pokémon-Opposite-Positive* and *Pokémon-Opposite-Negative* ($p = .001$), such that selecting the ‘Opposite’ cue when presented with a Pokémon and negative images resulted in that stimulus being evaluated positively ($M = 3.4, SE = .6$) whereas selecting the ‘Opposite’ cue when presented with a Pokémon and positive images resulted in that stimulus being evaluated negatively ($M = -3.7, SE = .6$).

Although self-reported ratings of the four Pokémon were all independently significant, the obtained effects were more pronounced when participants were required to affirm rather than negate the stimulus relation. In particular, positivity scores were larger when a Pokémon was trained as *Same-Positive* relative to *Opposite-Negative*, ($p = .001$), whereas negativity scores were larger when a Pokémon was trained as *Same-Negative*

relative to *Opposite-Positive*, ($p = .005$). In short, self-reported responding was not determined by simple stimulus pairings, but varied according to the relation established between stimuli by a contextual cue.

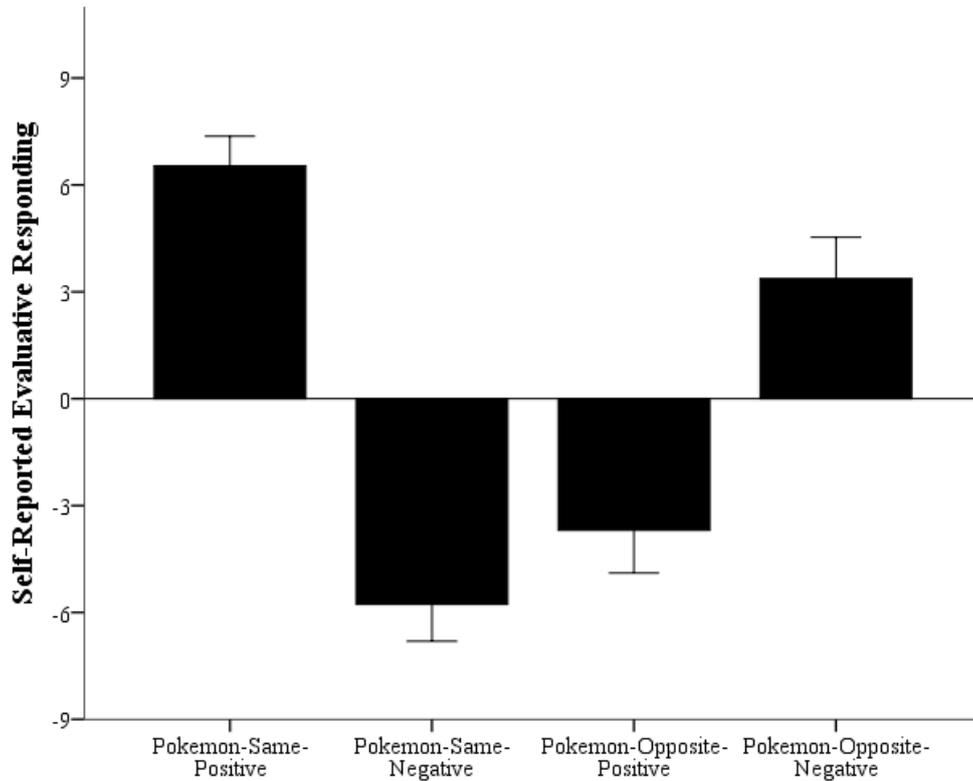


Figure 2.3. Mean likeability scores for each of the Pokémon characters as a function of the valenced images and contextual cue it was paired with. Error bars represent standard errors.

Indirect Procedure

Following the recommendations of Greenwald and colleagues (2003) response latency data was prepared using the D6 scoring algorithm. This transformation resulted in an IAT effect for each participant, reflecting the difference in mean response latency between consistent and inconsistent conditions divided by the overall variation in those latencies. For the ‘coordination’ IAT, scores were calculated so that positive values reflected a response bias for *Pokémon1-Same-Positive* relative to *Pokémon3-Same-Negative* while negative values indicated the reverse pattern of responding. At the same time, the ‘opposition’ IAT was scored so that negative values reflected a preference for *Pokémon4-Opposite-Negative* relative to *Pokémon2-Opposite-Positive* (see Figure 2.4). In-line with the assumptions of the

REC model, we predicted that brief and immediate relational responding towards the Pokémon characters would differ as a function of the contextual cue that governed the stimulus relation.

To test this assumption, IAT scores were submitted to a 2(IAT type: coordination vs. opposition) x 2(Block order; consistent vs. inconsistent) mixed-model ANOVA, with the latter factor manipulated between participants. A significant main effect was obtained for IAT type, $F(1, 73), = 53.1, p = .001, \eta^2_{\text{Partial}} = .42$, as well as a two-way interaction between IAT type and block order, $F(1, 73), = 13.4, p = .001, \eta^2_{\text{Partial}} = .16$. In order to specify this interaction, IAT effects for participants who completed the tasks in a consistent-first fashion were assessed separately from those who completed the tasks in an inconsistent-first manner. With respect to the consistent-first group, a significant main effect was observed for IAT type, $F(1, 41), = 93.7, p = .001, \eta^2_{\text{Partial}} = .70$, indicating that task performance was governed by stimulus relating rather than simple contiguity. Consistent with our predictions, performance on the coordination IAT was driven by a response bias favouring *Pokémon1-Same-Positive* relative to *Pokémon3-Same-Negative* ($M = .38, SE = .05$), $t(41) = 7.6, p = .001$. At the same time, and despite being directly paired with negative images, participants also favoured *Pokémon4-Opposite-Negative* relative to *Pokémon2-Opposite-Positive* on the opposition IAT ($M = -.36, SE = 0.6$), $t(41) = -6.3, p = .001$. A comparable pattern of responding was also evident for the inconsistent-first group, with a significant main effect obtained for IAT type, $F(1, 32), = 4.3, p = .05, \eta^2_{\text{Partial}} = .12$. Nevertheless, while participants displayed the expected response bias on the coordination IAT ($M = .20, SE = 0.7$), $t(32) = 2.9, p = .01$, behavioural effects were noticeably absent on the opposition IAT ($p = .6$). Taken together, the above findings suggest that brief and immediate relational responding was driven not simply by stimulus pairings but rather the relation established between stimuli by a contextual cue. Curiously, task performance was more robust when the first IAT block

participants encountered was consistent – rather than inconsistent - with what they had just learned.

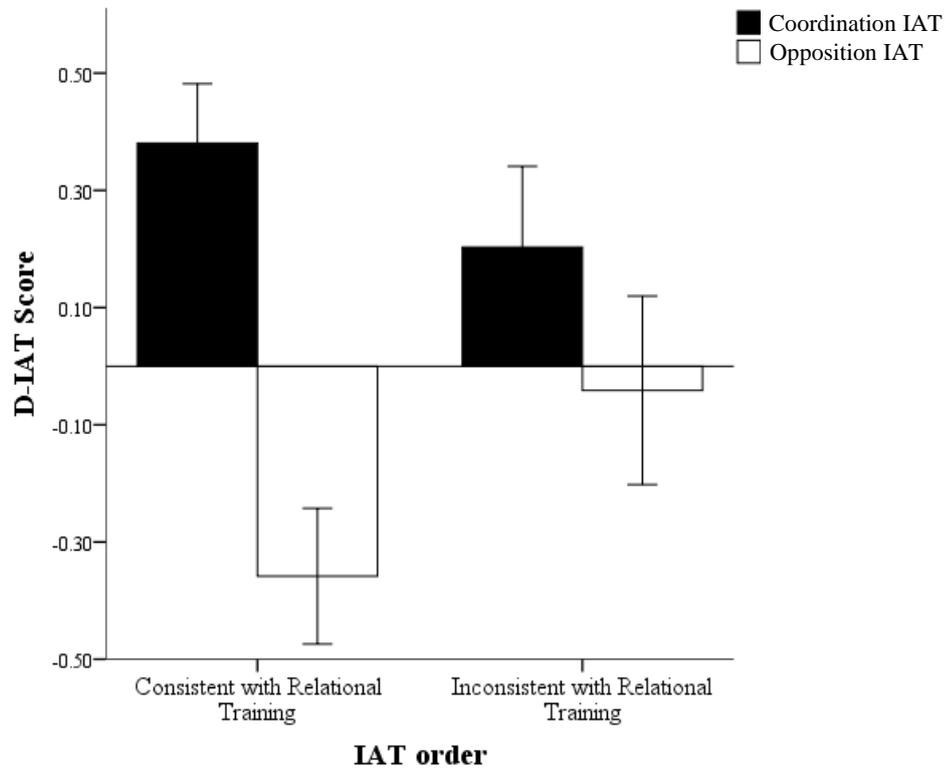


Figure 2.4. Mean D-IAT scores as a function of block order for both the coordination and opposition IATs. Error bars represent standard errors.

2.4 Discussion

A growing number of studies have shown contextual cues such as “friend/enemy”, “true/false” and “loves/loathes” can alter the psychological functions that are transformed through a stimulus relation. Such work raises an interesting question; how do cues such as words come to acquire their psychological functions and alter the functions of stimuli that they are related to? Drawing on RFT, the above study suggests that one answer lies in the behavioural phenomenon known as derived stimulus relating; once an individual has learned to arbitrarily relate one event with another under contextual control this history of learning can be brought to bear in an experimental situation - given appropriate cues to do so.

Consistent with this assumption, the results from Experiment 1 reveal that a stimulus can come to be liked or disliked – not simply based on the stimulus it is paired with - but

rather the relation established between those stimuli by a contextual cue. When a coordination ('Same') relation was established between a Pokémon character and valenced images the former was found to acquire the psychological function of the latter (e.g., *Pokémon-Same-Positive* was rated positively while the *Pokémon-Same-Negative* was rated negatively). Importantly, however, the direction of the obtained effect was completely reversed when selecting the 'Opposite' cue was reinforced in the presence of similar stimuli (e.g., participants rated *Pokémon-Opposite-Positive* as negative and *Pokémon-Opposite-Negative* as positive). This pattern of responding was not confined to self-reported ratings but was also observed on an indirect procedure (IAT) as well. For instance, when an IAT targeting the two coordination relations was administered participants displayed a clear response bias for the Pokémon that was *Same-Positive* relative to the Pokémon that was *Same-Negative*. Yet the direction of this effect was completely reversed when an IAT targeting the two opposition relations was administered (i.e., participants showed a response bias favouring the Pokémon that was *Opposite-Negative* over the Pokémon that was *Opposite-Positive*).

Although the current study is not the first to observe contextual control over evaluative responding it does provide an experimental analysis of how contextual cues come to acquire their psychological functions. In other words, while EC researchers have tended to select pre-existing cues from the individual's verbal repertoire (e.g., De Houwer, 2006; Förderer & Unkelbach, 2012; Peters & Gawronski, 2011) the learning history necessary to generate those cues in the first place is rarely, if ever, specified (although see Zanon et al., 2012). In contrast, the relational functions of 'Same' and 'Opposite' were established within the laboratory for two arbitrary symbols that had no prior meaning for the participant. This was achieved in-line with the predictions of RFT by providing a history of differential reinforcement involving training and testing across multiple exemplars. The success of this

strategy is evident in the contextually controlled patterns of responding observed on direct and indirect tasks as well as the labels applied to the cues by participants at the end of the experiment.

It should also be noted that verbal instructions about the stimulus relation were never dispensed nor was information about the intended meaning of the contextual cues provided. Rather relations between stimuli were established by providing a history of differential reinforcement for selecting a previously trained contextual cue in the presence of a Pokémon and positive or negative image. This strategy differs dramatically from recent studies that appeal to the verbal abilities of the individual without any reference to how language develops or influences stimulus relating (e.g., Fiedler & Unkelbach, 2011; Peters & Gawronski, 2011; Walther et al., 2011). To reiterate, we contend that relating stimuli under arbitrary contextual control is a type of generalized operant behaviour that is learned through multiple exemplar training which can give rise to various patterns of evaluative responding on both direct and indirect tasks.

Although contextual cue training established relational functions for the two symbols in-line with experimental expectations, a degree of behavioural variability was evident across participants, such that one cue was reported as meaning “same”, “similar” or “related” and the second as “opposite”, “different” or “dissimilar”. Critically, this variability (especially with respect to the ‘Opposite’ cue) may have subtle but important implications for contextually controlled evaluative responding. Consider, for example, a participant who relates a Pokémon as *Different-Positive* relative to a second participant who relates that Pokémon as *Opposite-Positive*. While it is unlikely that either participant will evaluate the Pokémon in a positive light, they may produce varied responses to the same stimulus. Put another way, whereas relating on the basis of a “different” cue involves responding to one event in terms of its relative difference from another along some specified dimension, the

magnitude of that difference is usually unspecified. Opposition relations, however, involve responding to one event in terms of its absolute difference from another along some specified continuum.

When applied to the current work, responding to a Pokémon as *Different-Positive* could have resulted in the Pokémon character being rated in a number of ways, from reduced positivity, to neutral or even negativity. In contrast, responding to a Pokémon as *Opposite-Positive* primarily entails negativity (although differences in absolute negativity are still possible). This variability in responding to the contextual cues suggests that our procedure needs to be refined so that the probability of opposition functions being abstracted from multiple exemplar training is increased relative to those involving difference. This task is complicated by the fact that all opposition relations entail difference but not all difference relations necessarily entail opposition. With this in mind, future work could ensure that any picture combinations involving relative differences along some specified dimension are replaced with those that clearly indicate opposition during contextual cue training (e.g., trials presenting pictures of night and day versus chalk and cheese).

It is also worth noting that the relational effects produced by the IAT were significantly more robust when participants completed the task in a manner that was consistent (relative to inconsistent) with previous training. Although the extraneous influence of block order has been well-documented with regard to pre-existing histories of learning (see Nosek, Greenwald & Banaji, 2007), the impact of this method factor on newly established histories has so far attracted little attention. Drawing on recent work by Ebert and colleagues (2009) it may be the case that the IAT not only assesses but actually modifies laboratory induced stimulus relations. In other words, and analogous to Heisenberg's uncertainty principle, the IAT may alter the probability of observing a particular response in the process of measuring it. To illustrate this more clearly, consider a participant who immediately

progresses from relational training to a “consistent-with-training” IAT block of trials. In this instance, the measurement context (i.e., the IAT) is arranged in such a way that participants have to respond in-line with a similar contingency that they recently encountered during training (e.g., relate Pokémon 1 with positive and Pokémon 3 with negative words). Thus in the process of testing the strength or probability of a particular relational response, the task may (inadvertently) further reinforce those stimulus relations.

In contrast, progression from relational training to an “inconsistent-with-training” block of trials results in an entirely different scenario. Participants contact a contingency that does not reinforce, but rather punishes, responding in accordance with recently established relations. Indeed, the person is required to relate in a manner that directly opposes everything they have just learned, and in doing so, the measure may serve to undermine the very history it is trying to capture. In other words, two very different IAT effects may be obtained depending on which block order participants are exposed to. On the one hand, completing the IAT in a consistent-first fashion may provide participants with an additional session of training before they actually have to respond inconsistently with what they have previously learned (i.e., the task may strengthen the resulting effect). On the other hand, beginning the IAT in an inconsistent-first fashion could partially undue the previously trained relations – thus weakening the resulting behavioural outcome. This explanation of task-order effects is in line with Ebert et al.’s (2009) findings and could account for why inconsistent-first IAT effects were diminished (rather than absent) in the above study. Future work could explore this possibility, and in particular, whether (a) newly established histories are more sensitive to the reversal in response contingencies than their well-established counterparts and (b) whether indirect tasks comprised of a block structure are more prone to these order effects relative to their counterparts that assess responding on a trial-by-trial basis.

Finally, while self-reported ratings of the four Pokémon characters were independently significant they did increase in magnitude when stimuli were related on the basis of coordination relative to opposition. For instance, the Pokémon trained as *Same-Positive* was rated more positively than its counterpart trained as *Opposite-Negative* while the Pokémon trained as *Same-Negative* was evaluated more negatively than the Pokémon trained as *Opposite-Positive*. One possible explanation, at least from an RFT perspective, concerns the differential response strength of coordination relative to opposition relating in everyday language use. According to this account, coordination relations - in which one event is identical, similar or the same as another - may represent the most fundamental and ubiquitous type of relational response emitting by verbally sophisticated humans (Hayes, et al., 2001). As such, it is possible that less complex and highly-practiced relational responses may give rise to stronger relational effects compared to other types of relations that are more complex and have been practiced to a lesser degree (e.g., opposition). While admittedly speculative, this assumption could be directly tested by exposing infants or developmentally disabled children without a history of coordination or opposition relating to equal amounts of training across both relations. Identifying human populations with limited verbal training and engineering those histories (as in current experiment) would shine a light on the role that the environment plays in developing these basic relational repertoires (see Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Luciano, Gómez-Becerra & Rodríguez-Valverde, 2007 for early work in this vein).

Chapter 3: A Derived Transformation of Functions through Combinatorially Entailed Relations as Measured by the IAT, IRAP and Self-Report Tasks

The results of Experiment 1 support the assumption that direct and indirect effects may be brought under relational rather than strictly associative control. Nevertheless, one potential drawback of this work is that only mutually entailed relations were established between the various stimuli, and in each case, psychological functions were established through direct reinforcement. While recognizing that stimuli can acquire functions via directly trained contingencies or stimulus co-occurrences, RFT makes the additional claim that novel functions may also be established and existing functions altered through the derivation of stimulus relations (i.e., the transformation of functions through not only mutual but combinatorially entailed relations). A convincing demonstration of this claim therefore requires a replication of the previous study with one important addition: a combinatorially entailed relation must be established involving two bi-directionally related stimuli separated by one intervening stimulus. If correct, then directly reinforcing *Pokémon 1-Same-Pokémon 2* and *Pokémon 2-Same-Pokémon 3* should lead to the spontaneous emergence of several derived relations without any explicit training or instruction to do so (e.g., *Pokémon 1-Same-Pokémon 3* and *Pokémon 3-Same-Pokémon 1*). In addition, if Pokémon 1 subsequently acquires an evaluative function, then the functions of Pokémon 2 and 3 should also be altered in-line with the relation established between those stimuli. In other words, a transformation of functions may represent the behavioural process by which humans come to like or dislike stimuli that have never been directly paired with appetitive or aversive events in the past. From this perspective, a stimulus can spontaneously acquire novel evaluative functions simply by participating in a derived relation with other (valenced) stimuli.

3.1 Experiment 2

With this in mind, Experiment 2 set out to test whether self-reported and IAT

performances could be governed by a transformation of functions through combinatorially entailed relations. The relational functions of ‘Same’ and ‘Opposite’ were once again established for two contextual cues using a broadly similar procedure as before. These cues were then used to form two coordination relations each comprised of three arbitrary Pokémon characters that shared no formal or physical properties with one another (*Pokémon1-Same-Pokémon2-Same-Pokémon3* and *Pokémon4-Same-Pokémon5-Same-Pokémon6*). Thereafter a relation of opposition was established between Pokémon 1 and negative images and Pokémon 4 and positive images (i.e., *Pokémon1-Opposite-Negative* and *Pokémon4-Opposite-Positive*). According to RFT this training should result in Pokémon 3 acquiring positive and Pokémon 6 negative functions despite the fact that neither stimulus was directly paired with valenced images or even indirectly paired with a stimulus that was itself paired with such images. At the same time, and consistent with Experiment 1, this derived transformation of functions should lead to a convergent pattern of responding on direct and indirect measures of evaluation.

Note that in Experiment 1 we used the IAT and self-report measures as relatively “indirect” benchmarks for the transformation of function through the trained relations. Indeed, unlike previously published studies, we never probed for the relational responses that were responsible for the obtained evaluative outcomes. To rectify this, Experiment 2 introduced a “derivation test” that sought to provide a more direct assessment of derived stimulus relating. Following relational training, an additional block of test trials were presented in which the final participant of either relation (Pokémon 3 or 6) was presented with a valenced image and the two contextual cues. Across a series of trials, participants were required to indicate whether a stimulus was *Same-Positive* or *Opposite-Negative* by selecting one of the two cues in the absence of corrective feedback.

If RFT is correct and responding towards the Pokémon characters is governed by

arbitrarily applicable relational responding then two distinct patterns of behaviour should be expected based on test performance. On the one hand, those participants who show a derived transformation of function (i.e., pass the derivation test) should not only self-report that Pokémon 3 is positive and Pokémon 6 is negative but also prefer Pokémon 3 over 6 when administered an IAT. On the other hand, those participants who fail to transform the functions of Pokémon 1 and 3 through the derived relations to Pokémon 4 and 6 (i.e., fail the derivation test) should show no preference for the latter stimuli on direct or indirect measurement procedures.

Finally, to ensure that exposure to the derivation test was not a necessary precondition for evaluative responding, the task was administered to half of the participants' immediately after training and to the other half at the end of the experiment. Counterbalancing exposure to this test allowed us to ensure that the obtained IAT and self-reported effects were not due to the direct pairing of Pokémon 3 or 6 with positive and negative images. Indeed, if stimulus pairings were responsible then evaluative responding should only emerge when the derivation test is administered before, but not after, the IAT and self-report measures. Furthermore, participants should also respond with ambivalence towards Pokémon 3 and 6 given that they were both paired an equal number of times with valenced images in the absence of any corrective feedback. However, if evaluative responding is the product of a common learning history (i.e., contextual cue and relational training) then IAT and self-reported ratings should be evident regardless of whether the test is encountered before or after their completion⁸. In short, Experiment 2 employed a contextually-controlled training procedure in an attempt to produce contextually-controlled derived stimulus relations in the form of contextually-

⁸ Testing for the formation of arbitrarily applicable relations typically involves a search for reflexivity, mutual, and combinatorial entailment. However, it is also possible to test for derivation using stimulus sorting (Smeets, Dymond, & Barnes-Holmes, 2000), response latency (Barnes-Holmes et al., 2005), or the transformation of function (Guinther & Dougher, 2010). That is, the transformation of function among stimuli that have not directly participated in a trained contingency (e.g., between A and C stimuli following A-B and B-C training) can be interpreted as evidence for an arbitrarily applicable relation between those stimuli.

controlled IAT and self-reported ratings. By omitting any direct pairings between Pokémon 1 and 3 or 4 and 6, the current study sought to highlight the central role that derived relations play in altering the psychological functions of indirectly related stimuli.

3.2 Method

Participants and Design

A total of sixty two undergraduate students (38 female), from 18 to 43 years ($M = 21.8$, $SD = 5.6$) were recruited from various departments at the National University of Ireland Maynooth and completed the study individually in exchange for a chocolate bar. The experiment consisted of a 2(*Task*: direct vs. indirect) x 2(*Derivation test*: pass vs. fail) x 2(*Derivation test time*: before vs. after) design, with the final factor manipulated between-participants. Two additional method factors were also manipulated; IAT block order and task presentation order (direct vs. indirect first). Data from eight individuals who failed to meet the mastery criteria during training were removed prior to analysis.

Materials

Stimuli. Twenty eight (non-participating) undergraduates were presented with a set of twenty four CS images (Pokémon) and asked to provide an evaluative rating for each using a scale ranging from -3 (Negative Feelings) to +3 (Positive Feelings) with 0 as the neutral point. The six Pokémon characters evaluated as the most neutral were selected as CSs (i.e., *desumasu*, *unown*, *yabukuron*, *klink*, *shibishirasu* and *dangoro*; *Mean rating* = .26, -.22, -.21, -.26, .00, -.14 respectively). The USs consisted of nine pleasant and nine unpleasant IAPS images. Mean evaluative ratings were 2.0 for the set of positive pictures and -2.5 for the set of negative pictures, $t(27) = 24.1$, $p = .001$.

Indirect procedure (IAT). Brief and immediate relational responding was assessed using an IAT. For each participant, Pokémon 3 and 6 served as one set of target labels and the words “Good” and “Bad” as another. Eight positively and negatively valenced adjectives,

matched in word length, valence extremity, and familiarity were employed as one set of attribute stimuli (positive: *happy, love, pleasure, fun, joy, kind, friendly, wonderful*; negative: *sick, horrible, disgusting, sad, terrible, awful, pain, vomit*) while four images of the two Pokémon characters at different orientations served as another set.

Procedure

After meeting the researcher and signing statements of informed consent, participants were seated in front of a computer and asked to pressed a button to initiate the experiment. The study consisted of four main stages: contextual cue training, relational training, a derivation test and (in)direct measures of evaluation.

Contextual Cue Training

The procedure used to establish the relational functions of ‘Same’ and ‘Opposite’ for two arbitrary symbols was broadly similar to that employed in Experiment 1.

Relational Training

Following contextual cue training two coordination relations - each comprising of three Pokémon characters – were generated via a protocol that was similar in many respects to that used in the previous study (*Pokémon1-Same-Pokémon2-Same-Pokémon3; Pokémon4-Same-Pokémon5-Same-Pokémon6*) (see Figure 3.1). The assignment of the six Pokémon characters within and between the two relations was randomly counterbalanced across participants. Training consisted of four separate phases, each comprised of a minimum of one and a maximum of three blocks of 50 trials.

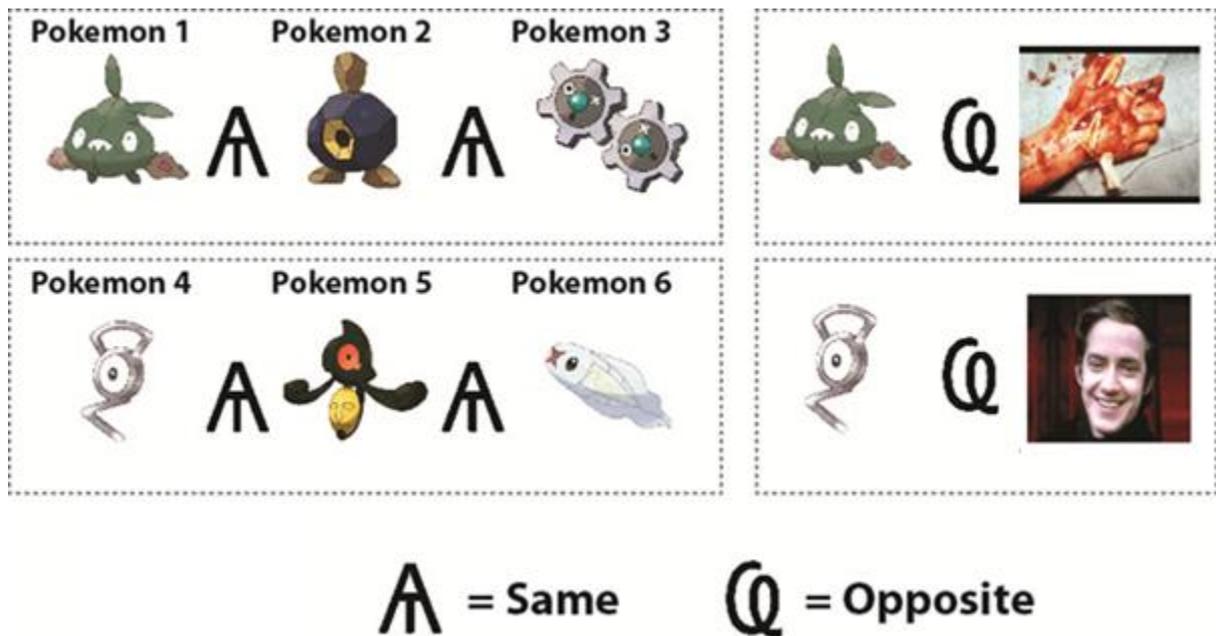


Figure 3.1. Schematic representation of the two experimentally established relational networks. In each case, three arbitrary Pokémon characters were first related to one another using the ‘Same’ cue (*Pokémon1- Pokémon2- Pokémon3* and *Pokémon4- Pokémon5- Pokémon6*). Thereafter, an opposition relation was established between the first member of either relation and positive or negative images (*Pokémon1-Opposite-Negative* and *Pokémon 4-Opposite-Positive*).

During the first phase, each trial displayed a Pokémon on the upper left and right sides of the screen as well as the two contextual cues at the bottom of the screen (see Figure 3.2). By differentially reinforcing the selection of one of the two contextual cues in the presence of a certain stimulus combination, four relations were trained (*Pokémon 1-Same-Pokémon2*; *Pokémon4-Same-Pokémon5*; *Pokémon1-Opposite-Pokémon5*; *Pokémon2-Opposite-Pokémon4*). For instance, when Pokémon 1 and 2 or Pokémon 4 and 5 were displayed together, selecting the ‘Same’ cue caused “Correct” to appear for 1000ms. Thereafter, all stimuli were cleared from the screen, followed by a 1000ms inter-trial interval (ITI) and the next trial. When an incorrect response was made, such as selecting ‘Opposite’ in the presence of the above stimuli, error feedback appeared in the middle of the screen and progression to the next trial was made contingent on selecting the correct (‘Same’) cue. Alternatively, when Pokémon 1 and 5, or Pokémon 2 and 4 were presented together, selecting the ‘Opposite’ cue

was reinforced and the ‘Same’ cue punished. Finally, in addition to the four stimulus relations, and to ensure that the ‘Same’ and ‘Opposite’ functions remained salient throughout the task, a number of contextual training trials were also interspersed within each block. The presentation of these trials and the assignment of Pokémon or contextual cues to the left and right sides of the screen was varied in a quasi-random order.

The second phase of training was identical to the first with the exception that four different relations were established (*Pokémon2-Same-Pokémon3*; *Pokémon5-Same-Pokémon6*; *Pokémon2-Opposite-Pokémon6*; *Pokémon3-Opposite-Pokémon5*). In the third phase participants were re-exposed to all eight stimulus relations trained during the previous two phases. Thereafter, and in the final phase, an evaluative function was established for the first stimulus in either relation by presenting it with a valenced image and requiring one of the two contextual cues to be selected. Specifically, either Pokémon 1 and a negative image, or Pokémon 4 and a positive image were presented at the top of the screen along with the two contextual cues at the bottom of the screen. In either case, choosing the ‘Opposite’ cue resulted in “Correct” being displayed, followed by the ITI and the next trial, while selecting the ‘Same’ cue produced error feedback. Throughout this section of the task, both Pokémon were presented with eight different images of the same valence (e.g., Pokémon 1 was always presented with negative images but never positive images and vice versa for Pokémon 4). To minimise the potential for a response bias, these critical CS-US trials were intermixed with a number of identity and contextual cue training trials that required the ‘Same’ cue to be selected.

To proceed from one phase to another, participants had to achieve a mastery criterion of 20 consecutively correct responses on a block of training trials followed by at least 20 out of 24 correct responses on a block of test trials. Failure to do so resulted in re-exposure to the

task; following a total of three training and testing blocks the participant was thanked, debriefed and their data discarded (eight participants were removed on this basis).

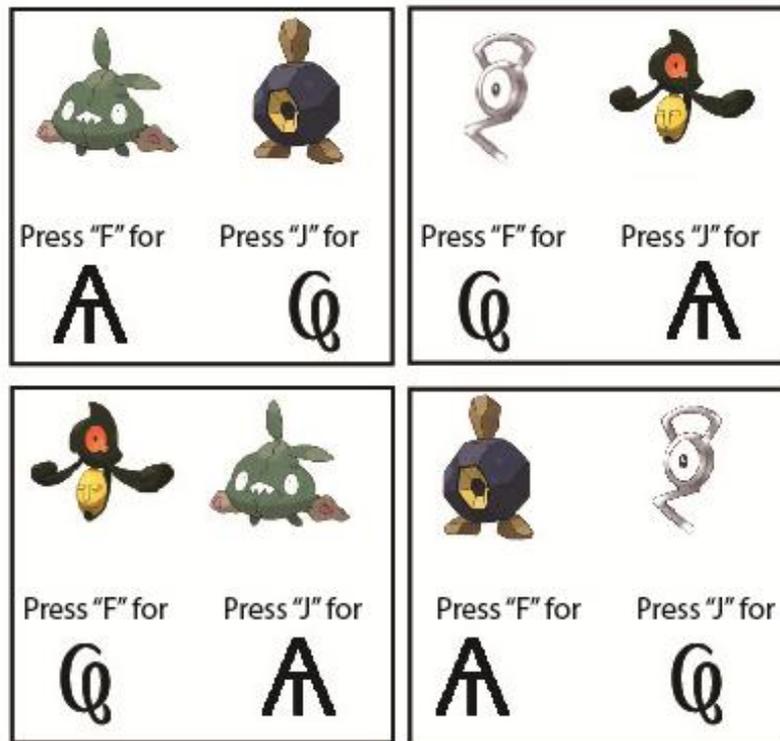


Figure 3.2. Examples of the four trials involved in the first phase of relational training. Each trial consisted of two Pokémon at the top of the screen and the two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being displayed in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).

Test for Derived Relational Responding

In order to determine whether the two relational networks were formed as predicted (and the functions established for Pokémon 1 and 4 were transformed through those relations to Pokémon 3 and 6) participants were exposed to a derivation test. On each trial, Pokémon 3 or 6 and a positive or negative image were presented at the top of the screen and the two contextual cues were displayed at the bottom. Participants were asked to “*click on the symbol that describes the relationship between the two pictures at the top of the screen*”. They were also provided with an option to report that “*I do not know*” the relationship between the stimuli. Clicking on a contextual cue (or the “*I don’t know*” button) removed all stimuli from the screen, followed by an inter-trial interval of 500ms and the next trial. Testing involved a

block of twelve trials, six of which presented Pokémon 3 with a positive or negative image while the remaining six presented Pokémon 6 with a positive or negative image. No feedback was provided for any response emitted during this task.

If a transformation of functions through the derived relations occurred as predicted, then participants should consistently select the ‘Opposite’ cue when shown Pokémon 3 and a negative image; the ‘Same’ cue given Pokémon 3 and a positive image; the ‘Opposite’ cue when presented with Pokémon 6 and a positive image; and the ‘Same’ cue given Pokémon 6 and a negative image. Stated more precisely, participants who produced a minimum of 10 out of 12 correct responses were defined as passing the test while those who failed to do so were defined as having failed the task.

Indirect Procedure (IAT)

Following training participants completed direct and indirect measures of evaluative responding. To assess whether a derived transformation of functions occurred, brief and immediate relational responding towards Pokémon 3 relative to Pokémon 6 was assessed using a similar IAT as before. Once again, if the relational networks were established in-line with prior training then response latencies should be shorter during consistent relative to inconsistent IAT trials. Specifically, faster responding should be obtained when participants have to relate Pokémon 3 with positive relative to negative stimuli and Pokémon 6 with negative relative to positive stimuli.

Direct Procedures

Self-reported ratings of the Pokémon, contextual cue meaning and demand compliance tasks were similar to those employed in Experiment 1.

3.3 Results

Data Preparation

Contextual cue meaning. Of the current sample fifty two participants (96%) reported

the relational functions of the contextual cues in-line with experimental expectations. On the one hand, thirty two participants (59%) rated the ‘Same’ cue as meaning “same”; fourteen participants (26%) rated it as meaning “similar” while the remaining six participants (11%) used terms such as “alike”, or “connected”. On the other hand, twenty nine participants (54%) rated the ‘Opposite’ cue as “opposite”; eighteen (33%) rated it as meaning “different” while the remaining five participants (9%) used a variety of terms such as “difference”, “not compatible” or “dissimilar”. The data for the two participants who reported incorrect relational functions were removed prior to analysis. Reanalyzing the data with these participants included did not change any of the self-reported or IAT effects reported below.

Demand compliance. Three of the fifty two participants reported that they intentionally responded to the Pokémon in-line with the presumed expectations of the experimenter. Once again, their data were excluded (reanalyzed the data with these data-points included did not influence any of the statistical conclusions drawn below).

Preliminary analyses. Counterbalancing the assignment of Pokémon within and across coordination relations, as well as the order of direct or indirect tasks had no effect, so analyses were collapsed across these two factors. Half of the participants failed while the other half passed the derivation test.

Self-Reported Ratings

Mean evaluative ratings as a function of derivation test performance are presented in Figure 3.3. As seen in the graph, the direction and magnitude of evaluative responding was contingent on the opposition relation established between Pokémon 1 or 4 and valenced images. For instance, reinforcing the selection of the ‘Opposite’ cue in the presence of Pokémon 1 and negative images resulted in that stimulus being rated positively, whereas selecting the ‘Opposite’ cue when presented with Pokémon 4 and positive images resulted in that stimulus being rated negatively. Building on our previous findings, these evaluative

functions were transformed through the coordination relations to other related stimuli, such that Pokémon 2 and 3 were also rated positively while Pokémon 5 and 6 were rated negatively.

In order to test these assumptions - and in particular - whether a derived transformation of functions from *Pokémon1-Opposite-Negative* to Pokémon 3 occurred, self-reported ratings for Pokémon 1, 2 and 3 were submitted to a $3(\text{Pokémon}) \times 2(\text{Derivation test performance: pass vs. fail}) \times 2(\text{Derivation test time: before vs. after})$ mixed-model ANOVA. A significant main effect emerged for derivation performance, $F(1, 45) = 23.6, p = .001, \eta^2_{\text{Partial}} = .34$, indicating that evaluative responding significantly differed depending on whether the participant derived the stimulus relations (no main or interaction effects emerged for test time; all $ps > .3$). Follow-up analyses revealed that those who passed the derivation test rated Pokémon 1, which was directly paired with negative images, as significantly more positive than those who failed, $F(1, 48) = 26.3, p = .001, \eta^2_{\text{Partial}} = .36, (Ms = 2.5 \text{ vs. } -.1, \text{ respectively})$. Likewise, and despite the fact that they were never paired with valenced images, Pokémon 2, $F(1, 48) = 22.7, p = .001, \eta^2_{\text{Partial}} = .33, (Ms = 2.4 \text{ vs. } -.1)$ and Pokémon 3 were also rated as significantly more positive by those who passed the test relative to those who failed, $F(1, 48) = 11.6, p = .001, \eta^2_{\text{Partial}} = .20, (Ms = 1.9 \text{ vs. } -.1)$. Indeed, while ratings for each of the Pokémon were significantly different from neutral for the pass group (all $ps < .001$) they were not for the fail group (all $ps > .3$).

When a similar set of analyses were conducted for the *Opposite-Positive* relation (i.e., Pokémon 4, 5 and 6), a main effect emerged for derivation performance, $F(1, 45) = 42.6, p = .001, \eta^2_{\text{Partial}} = .49$, while no main or interaction effects emerged for test time (all $ps > .4$). Once again, those who passed the test rated Pokémon 4 (i.e., the stimulus directly paired with positive images) as significantly more negative than those who failed, $F(1, 48) = 40.1, p = .001, \eta^2_{\text{Partial}} = .46, (Ms = -2.8 \text{ vs. } .2)$. Moreover, and despite the fact that they were never

paired with valenced images, Pokémon 5, $F(1, 48) = 27.5, p = .001, \eta^2_{Partial} = .37, (Ms = -2.4$ vs. $.04)$ and Pokémon 6, $F(1, 48) = 26.8, p = .001, \eta^2_{Partial} = .36, (Ms = -2.3$ vs. $.3)$ were also rated as significantly more negative by those who passed the test relative to those who failed. Although ratings for each of the Pokémon were significantly different from neutral for the pass group (all $ps < .001$), no effect reached significance for the fail group (all $ps < .2$). Overall, evaluative responding was under relational rather than simple associative control, such that Pokémon were liked or disliked depending on what contextual cue governed the stimulus relation. These functions were also transformed through the derived relation to stimuli that were never directly paired with or reinforced in the presence of valenced images.

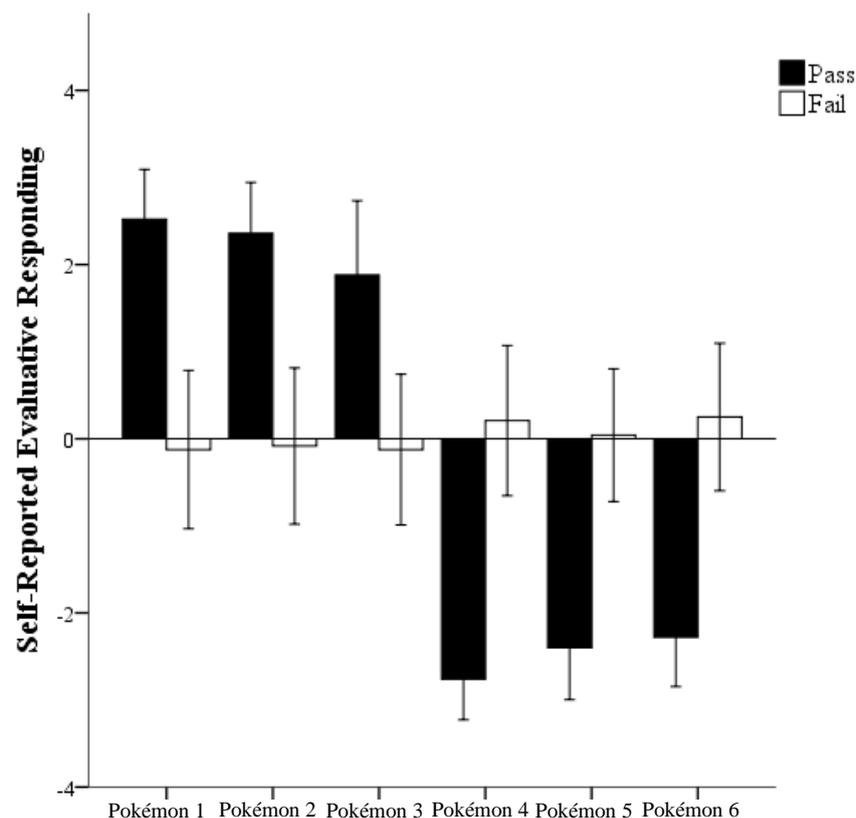


Figure 3.3. Mean likeability scores for each of the Pokémon characters as a function of derivation test performance (pass vs. fail). Error bars represent stand errors.

Indirect Procedure (IAT)

Data were transformed using the D6 algorithm and scored so that positive values

reflected a response bias for Pokémon 3 relative to 6 while negative values indicated an opposite pattern of responding (see Figure 3.4). When submitted to a 2(*Derivation test time*) x 2(*Derivation test performance*) x 2(*IAT block order*) mixed-model ANOVA, main effects emerged for test performance, $F(1, 48) = 13.9, p = .001, \eta^2_{\text{Partial}} = .25$, and IAT block order, $F(1, 48) = 14.4, p = .001, \eta^2_{\text{Partial}} = .26$, in addition to a three-way interaction between test performance, test time and block order, $F(1, 48) = 5.5, p = .02, \eta^2_{\text{Partial}} = .12$ (note that in the interest of clarity Figure 3.4 does not include block order as a variable).

To specify the obtained three-way interaction, the effects of block order and test time were assessed separately for those who passed versus failed the derivation test. With respect to the pass group, a one-way ANOVA indicated that encountering the test before or after the evaluative measures did not influence responding towards the Pokémon characters, $F(1, 24) = 1.6, p = .2$, nor did the order in which the IAT was completed, $F(1, 24) = 1.9, p = .2$. As predicted, a response bias emerged for Pokémon 3 relative to Pokémon 6 in-line with prior relational training ($M = .3, SE = .08$), $t(24) = 3.7, p = .001$. In direct contrast to these findings, participants who failed the derivation test showed a significant main effect for IAT order, $F(1, 23) = 18.9, p = .001, \eta^2_{\text{Partial}} = .49$ as well as a significant two-way interaction between IAT order and test time, $F(1, 23) = 7.1, p = .02, \eta^2_{\text{Partial}} = .26$. In other words, when the fail group encountered an inconsistent-first IAT they responded in a manner that was inconsistent with prior training - regardless of whether the derivation test was encountered before ($M = -.49, SD = .23$) or after that IAT ($M = -.32, SD = .45$). However, when the fail group encountered a consistent-first IAT they responded in a manner that was consistent with prior training and these effects were larger when the derivation test was administered before ($M = .47, SD = .19$) rather than after the IAT ($M = -.09, SD = .26$).

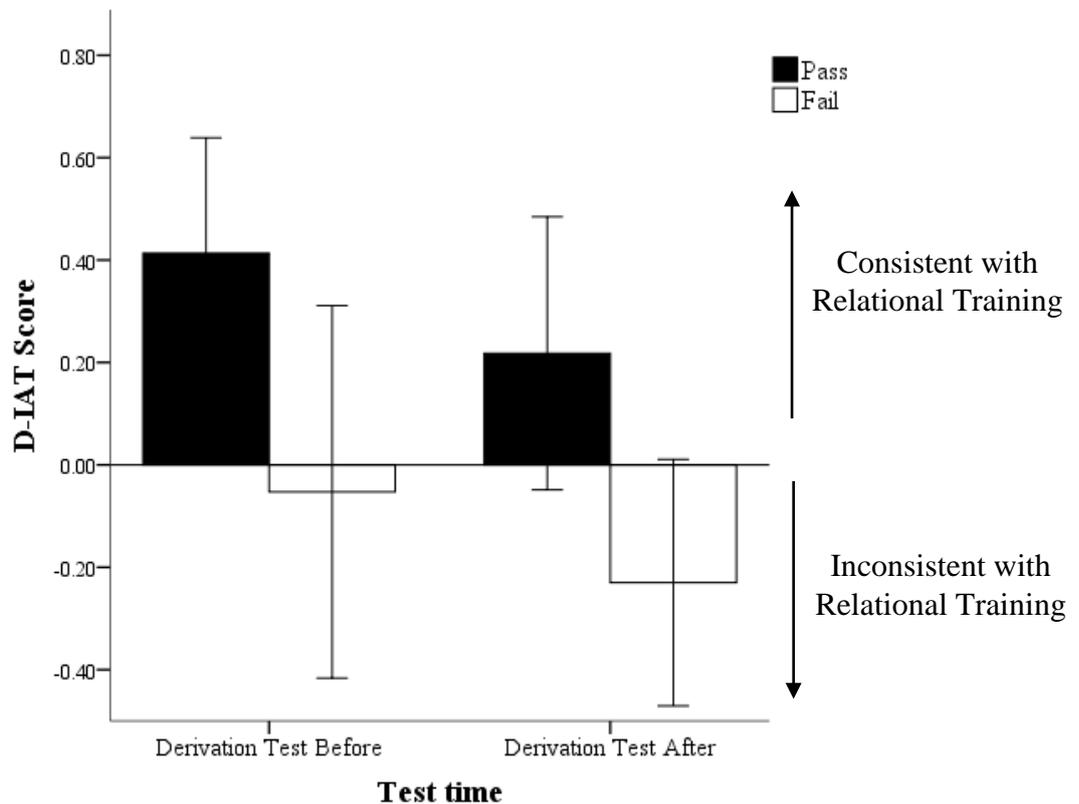


Figure 3.4. Mean D-IAT scores as a function of derivation test performance (pass vs. fail) and derivation test time (before vs. after). Error bars represent standard errors.

Discussion

The above findings not only replicate Experiment 1 (in so far as direct and indirect task performance was brought under arbitrarily applicable contextual control) but extend those findings to combinatorially entailed relations. Consistent with the previous study, participants rated Pokémon 1 as positive once an opposition relation was established between that stimulus and negative images. Perhaps more interesting, however, is that when Pokémon 1 participated in a coordination relation with Pokémon 2 and 3, the latter two stimuli were also rated positively despite the fact that (a) they were never paired with valenced images during the task and (b) the functions of Pokémon 1 were established *after* the coordination relation was generated. At the same time, a converse pattern of responding was observed when the ‘Opposite’ cue was reinforced in the presence of Pokémon 4 and positive images. Generating a coordination relation between Pokémon 4, 5 and 6 resulted in participants not

only rating Pokémon 4 as negative but evaluating the other two stimuli in a similar fashion. These findings applied across both direct and indirect procedures alike, with participants demonstrating a response bias for Pokémon 3 relative to 6 when an IAT was administered.

Drawing on Relational Frame Theory, we expected that evaluative responding would depend on whether participants passed a derivation test. This prediction was clearly confirmed in the current study with contextually-controlled patterns of evaluative responding evident only when participants passed that test. Those who failed to do so rated each of the Pokémon characters as neutral while their IAT performance was largely determined by whether they completed the task in a consistent or inconsistent-first order.

3.4 Experiment 3

Although these results offer support for a relational interpretation of evaluative responding it should be noted that Experiments 1 and 2 limited their analyses to a single indirect measure (the IAT). Although unlikely, it is possible that these experimentally induced IAT effects reflect task-specific properties of the measure rather than a general and overarching behavioural process as claimed (e.g., Deutsch & Gawronski, 2009). Therefore the goal of Experiment 3 was to investigate whether similar findings would also emerge when an alternative indirect procedure (the IRAP) was used in place of the IAT. Adopting a new measure not only increases the generalisability of the previous findings but protects against method-specific artifacts associated with any one task. The IRAP also allowed us to address specific questions concerning the directionality of brief and immediate relational responding. Given the relativistic nature of the IAT it has not been possible to identify whether task outcomes were driven by a “preference” or a “dislike” for one stimulus over another. In contrast, by presenting specific label and target stimuli together on a certain trial and requiring a particular response to be emitted quickly and accurately, the IRAP can target four separate stimulus relations independently from one another. In doing so, the degree to which

brief and immediate relational responding is driven by coordination relations (*Pokémon3-Same-Good; Pokémon6-Same-Bad*), opposition relations (*Pokémon3-Opposite-Bad; Pokémon6-Opposite-Good*) or some combination of the two can be determined.

3.5 Method

Participants and Design

Fifty seven undergraduate students (39 women) ranging from 18 to 34 years ($M = 21.5$, $SD = 4.5$) were drawn from the same research pool, with the constraint that they had not taken part in the previous study. This experiment consisted of a $2(\textit{Derivation test performance}) \times 2(\textit{Task order}) \times 2(\textit{IRAP block order})$ factorial design with the latter two factors manipulated between-subjects. The data from several participants were omitted from analyses due to failures to attain either the mastery criteria during training ($n = 11$) or maintain test performance criteria during the IRAP ($n = 7$). Therefore the final sample was comprised of a total of thirty nine participants.

Procedure

A similar set of procedures were administered as in the previous study with two notable exceptions. First, given that exposure to the derivation test was not found to influence evaluative responding, all participants received that test prior to completing the various measures of evaluation. Second, all participants completed an IRAP in place of the IAT.

Indirect Procedure (IRAP)

The IRAP operates on the assumption that participants should respond with greater speed and accuracy when they have to relate stimuli in a manner deemed consistent (relative to inconsistent) with their prior history of learning. The difference in time taken to respond across alternating blocks of trials - defined as the IRAP effect - is assumed to provide an index of the strength or probability of the targeted relations.

Overall, the task consists of a minimum of two practice followed by six test blocks,

each comprised of 24 trials. Although the same response contingency applied to all the trials within a block, these contingencies were reversed across successive blocks, with participants exposed to an alternating sequence of history consistent versus inconsistent response contingencies. On each trial, four stimuli were simultaneously presented on-screen: one of two label stimuli (either Pokémon 3 or 6) was presented at the top centre, a target stimulus (positive or negative adjective) directly below in the middle and two response options (“True” and “False”) at the bottom left and right corners of the screen (see Figure 3.5). The instructions “Press ‘D’ for” and “Press ‘K’ for” appeared above the left and right response options respectively.

During a block of consistent trials, participants were required to respond in accordance with the previously trained stimulus relations (*Pokémon3-Good-True; Pokémon3-Bad-False; Pokémon6-Bad-True; Pokémon6-Good-False*). For instance, when presented with Pokémon 3 and a positive word, or Pokémon 6 and a negative word, selecting “True” cleared all stimuli from the screen for 400ms followed by the onset of the next trial. If an inconsistent response was emitted, such as selecting “False” when presented with Pokémon 3 and a positive word or Pokémon 6 and a negative word, a red “X” appeared in the middle of the screen. To remove the error feedback and progress to the ITI, participants were required to emit the correct response. If a participant failed to respond within 2000ms from the start of a trial the words “Too Slow” appeared below the target stimulus and remained there until one of the response options was selected. During a block of inconsistent trials, responding in a manner that was incongruent with the previously trained relations was reinforced (*Pokémon3-Good-False; Pokémon3-Bad-True; Pokémon6-Bad-False; Pokémon6-Good-True*). Exposure to either the consistent or inconsistent first sequence of blocks was randomly counterbalanced across participants while the left-right positioning of the two response options and presentation of a particular trial-type was varied quasi-randomly within each block of trials.



Figure 3.5. Examples of the four trial-types used in the IRAP. A label stimulus (Pokémon 3 or 6), target word (nasty, nice, disgusting, pleasant, etc.), and two response options (True and False) appeared simultaneously on each trial. Selecting the option deemed correct on any given trial resulted in “Correct” being displayed in the middle of the screen while selecting the option deemed incorrect caused “Incorrect” to appear.

To make the participant’s history of learning with respect to Pokémon 3 and 6 apparent, the task required that they respond with both speed (e.g., median response time of less than 2,000ms) and accuracy (at least 80% correct responses). During the practice phase, the IRAP provided participants with up to four opportunities to achieve these two criteria (i.e., a total of eight practice blocks). Achieving the mastery criteria resulted in exposure to a series of six test blocks while failure to do so resulted in participants being thanked, debriefed and their data discarded. The procedure for the test blocks was similar to that of the practice, with the exception that no performance criteria were required to progress from one block to the next. On-screen instructions informed participants that “*this is a test*” and to “*go fast*” while trying to be as accurate as possible. Following the completion of all six blocks, the participant was thanked and debriefed. Failure to maintain speed and accuracy criteria across

each of the test blocks resulted in the participant's data being discarded prior to analysis (seven participants were removed on this basis).

If the relational networks were established in-line with prior training then response latencies should be shorter during consistent relative to inconsistent IRAP trials. Specifically, faster responding should be obtained when participants have to relate *Pokémon 3-Positive-True* and *Pokémon 3-Negative-False* relative to *Pokémon 3-Negative-True* and *Pokémon 3-Positive-False*. Likewise, they should also respond with greater speed when they have to relate *Pokémon 6-Negative-True* and *Pokémon 6-Positive-False* relative to *Pokémon 6-Positive-True* and *Pokémon 6-Negative-False*.

3.6 Results

Data Preparation

Contextual cue meaning. Of the current sample thirty eight participants (97%) reported the relational functions of the contextual cues in-line with experimental expectations. On the one hand, twenty two participants (58%) rated the 'Same' cue as meaning "same"; twelve participants (32%) rated it as meaning "similar" while the remaining three participants (10%) used terms such as "alike", or "connected". On the other hand, twenty participants (53%) rated the 'Opposite' cue as "opposite"; eleven (29%) rated it as meaning "different" while the remaining seven participants (18%) used a variety of terms such as "unlike", "not same" or "not connected". The data for the single participant who reported incorrect relational functions were removed prior to analysis. Reanalyzing the data with this participant included did not influence any of the statistical conclusions drawn below.

Demand compliance. None of the participants who completed the study reported that they intentionally responded to the stimuli in-line with the presumed expectations of the experimenter.

Preliminary analyses. Counterbalancing the order of direct and indirect procedures as well as IRAP test blocks had no significant effect, so analyses were collapsed across these two factors. Eighteen participants (46%) failed the derivation test while twenty one (54%) passed the test.

Self-Reported Ratings

Mean evaluative responses for the Pokémon characters as a function of derivation test performance are presented in Figure 3.6. To investigate whether a derived transformation of functions from *Pokémon1-Opposite-Negative* to Pokémon 3 occurred, ratings for Pokémon 1, 2 and 3 were submitted to a 3(*Pokémon*) x 2(*Derivation test performance*) mixed-model ANOVA. Analyses revealed a significant main effect for derivation performance, $F(1, 36) = 18.6, p = .001, \eta^2_{\text{Partial}} = .34$ with evaluative responding towards the Pokémon differing depending on whether participants passed or failed the test. Consistent with Experiment 2, those who passed the derivation test rated the first, $F(1, 37) = 29.9, p = .001, \eta^2_{\text{Partial}} = .45$, ($M_s = 2.3$ vs. $-.4$, respectively) second, $F(1, 37) = 32.5, p = .001, \eta^2_{\text{Partial}} = .48$, ($M_s = 2.4$ vs. 0) and third Pokémon more positively than those who failed, $F(1, 37) = 22.4, p = .001, \eta^2_{\text{Partial}} = .38$, ($M_s = 2.7$ vs. $.5$). Although ratings significantly differed from zero for the pass group (all $p_s < .001$) no effect reached significance for the fail group (all $p_s > .06$). Finally, and unlike the previous study, a main effect also emerged for Pokémon, $F(2, 36) = 4.8, p = .01, \eta^2_{\text{Partial}} = .12$, with ratings differing depending on a stimulus' location in the relation. Post-hoc tests indicated that Pokémon 1 was evaluated less positively than Pokémon 3 ($p = .04$) while Pokémon 2 was marginally less positive than Pokémon 3 ($p = .08$) (no difference emerged between Pokémon 1 and 2; $p = .8$).

When a similar set of analyses were conducted for the *Opposite-Positive* relation (i.e., Pokémon 4, 5 and 6) a significant main effect for test performance emerged, $F(1, 36) = 18.6, p = .001, \eta^2_{\text{Partial}} = .34$. Once again, participants who passed the derivation test rated the first,

$F(1, 37) = 9.8, p < .01, \eta^2_{Partial} = .22, (Ms = -2.5 \text{ vs. } -.8, \text{ respectively}),$ second, $F(1, 37) = 15.9, p < .01, \eta^2_{Partial} = .31, (Ms = -2.0 \text{ vs. } -.3)$ and third Pokémon as significantly more negative relative to those who failed that test, $F(1, 37) = 23.2, p < .001, \eta^2_{Partial} = .39, (Ms = -2.2 \text{ vs. } -.1).$ While ratings for each of the Pokémon were significantly different from zero for the pass group (all $ps < .001$) no effect reached significance for the fail group (all $ps > .1$). Interestingly, a main effect was once again obtained for Pokémon, $F(2, 36) = 4.8, p = .01, \eta^2_{Partial} = .12,$ with post-hoc tests indicating that Pokémon 4 was significantly more negative than Pokémon 5 ($p = .04$), and marginally more so than Pokémon 6 ($p = .08$). When taken together, these results replicate those of Experiment 2 and reveal that only those participants who successfully derived the relation between stimuli showed evidence of self-reported evaluative responding.

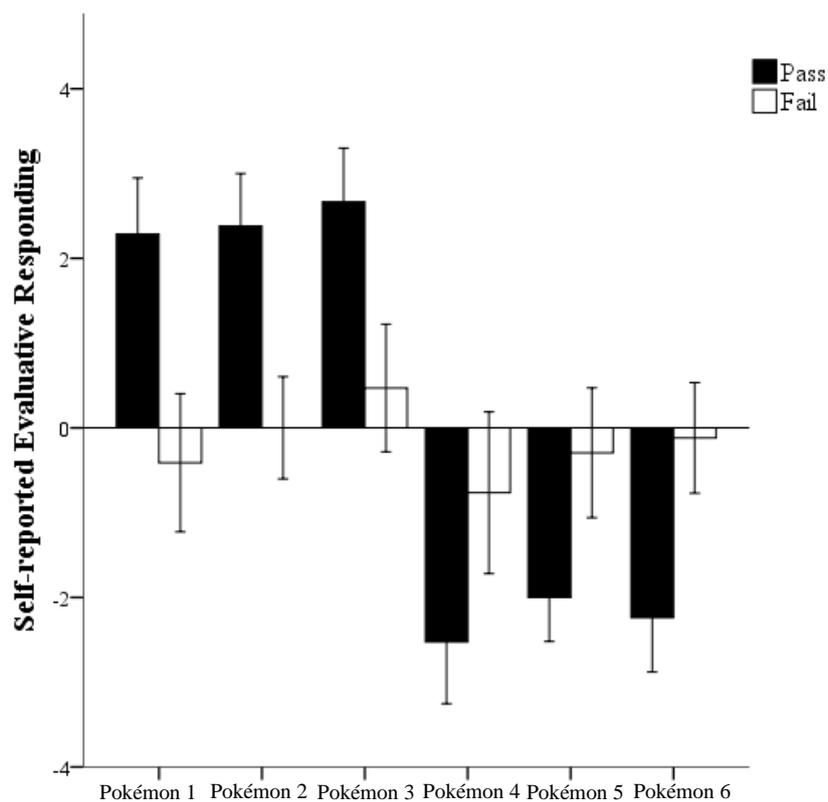


Figure 3.6. Mean likeability scores for each of the Pokémon characters as a function of derivation test performance (pass vs. fail). Error bars represent standard errors.

Indirect Procedure (IRAP)

The primary datum for the IRAP was response latency, defined as time in milliseconds from the onset of a test trial until a correct response was emitted. An adaptation of Greenwald, Nosek, and Banaji's (2003) D-algorithm was used to transform latencies from the six test blocks into four D_{IRAP} effect size scores, one for each of the stimulus relations assessed by the task (for a full description of the D_{IRAP} transformation see Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009). In the current study, these relations (termed IRAP trial-types) were *Pokémon 3-Good*, *Pokémon 6-Good*, *Pokémon 3-Bad* and *Pokémon 6-Bad*. Scores were calculated so that positive values indicated a response pattern that was consistent with prior relational training and negative values an opposite pattern of responding (see Figure 3.7).

When submitted to a 4(*Trial-types*) x 2(*Derivation test performance*) mixed-model ANOVA, a significant main effect was obtained for IRAP trial-type, $F(3, 36) = 20.6, p = .001, \eta^2_{\text{Partial}} = .36$, and test performance, $F(1, 36) = 4.0, p = .05, \eta^2_{\text{Partial}} = .1$, as well as a significant two-way interaction between these factors, $F(3, 36) = 3.2, p = .03, \eta^2_{\text{Partial}} = .1$. In order to specify this interaction, trial-type data for those who passed and failed the derivation test was assessed separately.

A significant main effect of trial-type emerged for participants who passed the test, $F(3, 20) = 12.1, p = .001, \eta^2_{\text{Partial}} = .38$. A series of one sample t-tests revealed a significant effect for the *Pokémon 6-Bad*, ($M = .26, SD = .44$), $t(20) = 2.8, p = .01$, and *Pokémon 3-Good* trial-types, ($M = .60, SD = .33$), $t(20) = 8.3, p = .001$, consistent with a derived transformation of functions through the stimulus relations. Nevertheless, comparable effects were notably absent when the same participants were required to negate the derived relations (i.e., *Pokémon 6-Good* and *Pokémon 3-Bad*; both $ps > .3$). Although participants who failed the derivation test also demonstrated a significant effect for trial-type, $F(3, 16) = 11.8, p = .001$,

$\eta^2_{\text{Partial}} = .4$, they responded in a manner that was entirely incoherent with prior training; endorsing the *Pokémon 6-Good*, ($M = -.37$, $SD = .35$), $t(16) = -4.6$, $p = .001$ and *Pokémon 3-Good* trial-types, ($M = .26$, $SD = .42$), $t(16) = 2.5$, $p = .02$, while showing no effect for either *Pokémon 6-Bad* ($p = .2$), or *Pokémon 3-Bad* ($p = .1$).

Finally, an overall mean D_{IRAP} score was calculated for each participant by subtracting the latencies in the consistent blocks from the inconsistent blocks across all four IRAP trial-types. A positive score indicates an evaluative bias in-line with relational training (i.e., preference for Pokémon 3 relative to Pokémon 6), whereas a negative score reflects a reverse pattern of responding. Participants who passed the derivation test showed a moderate D_{IRAP} effect ($M = .24$, $SE = .06$) whereas those who failed demonstrated no effect at all ($M = .05$, $SE = .06$), with a significant difference emerging between the two groups, $F(1, 37) = 4.0$, $p < .05$, $\eta^2_{\text{Partial}} = .1$.

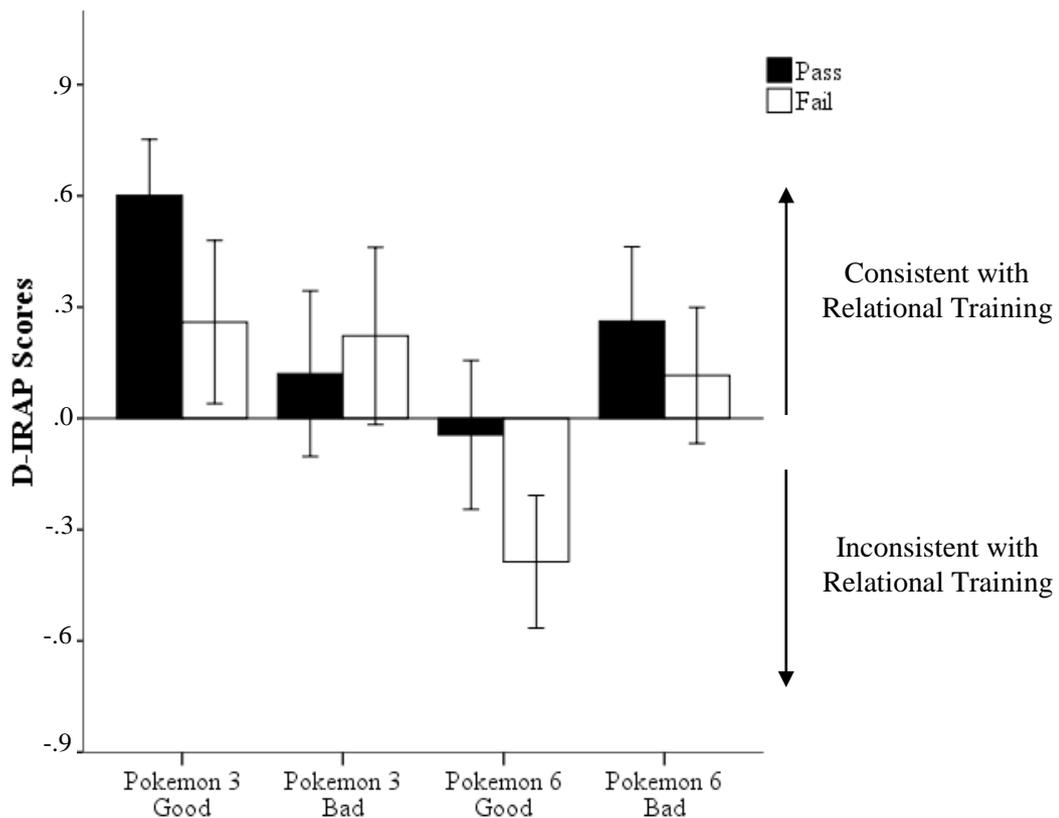


Figure 3.7. Mean D_{IRAP} scores for each of the four trial-types as a function of derivation test performance (pass vs. fail). A positive score indicates responding in-accordance with prior

training while a negative score indicates responses that are inconsistent with training. Error bars represent standard errors.

Discussion

The results from Experiment 3 further highlight that a derived transformation of functions through combinatorially entailed relations can lead participants to like or dislike stimuli that have never been directly paired with appetitive or aversive events in the past. When a coordination relation was established between Pokémon 1, 2 and 3 (followed by an opposition relation between Pokémon 1 and negative images) participants responded positively to all three stimuli in that relation. Likewise, when a second coordination relation was formed between Pokémon 4, 5 and 6 (followed by an opposition relation between Pokémon 4 and positive images) participants responded negatively to each of the stimuli within that relation. Critically, however, these patterns of responding were only observed when participants passed the derivation test. Those that failed to do so produced no evidence of evaluative responding towards any the Pokémon characters despite the fact that they successfully passed all stages of contextual cue and relational training.

A similar set of findings emerged when a novel indirect procedure (the IRAP) was used to assess brief and immediate relational responding. The pass group quickly and accurately related Pokémon 3 with positive words as well as Pokémon 6 with negative words, indicating that derived relations were formed as predicted. However, comparable effects were absent when those same participants were required to respond to *Pokémon6-Good* or *Pokémon3-Bad* as false. Put another way, significant IRAP effects emerged only when participants affirmed the derived relations, but not when they were required to negate those same relations. Once again, participants who failed the derivation test responded in an incoherent fashion across the majority of IRAP trial-types; incorrectly responding to

Pokémon6-Good as true while showing no effects for either *Pokémon3-Bad* or *Pokémon6-Bad*.

3.7 General Discussion

The current research examined whether stimuli could spontaneously acquire novel psychological functions by participating in derived relations with other valenced stimuli. The results from Experiments 2 and 3 provide clear support for this claim, such that evaluative responding was only evident when participants passed the derivation test. For instance, when two coordination relations were established (*Pokémon1-Same-Pokémon2-Same-Pokémon3* and *Pokémon4-Same-Pokémon5-Same-Pokémon6*) followed by an opposition relation between Pokémon 1 and negative images as well as Pokémon 4 and positive images, the pass group rated Pokémon 1, 2 and 3 positively and Pokémon 4, 5 and 6 negatively. Participants also responded with a high degree of speed and accuracy when Pokémon 3 had to be categorized (IAT) or related (IRAP) with positive words and Pokémon 6 with negative words. These direct and indirect effects are particularly noteworthy given that neither Pokémon 3 nor 6 was ever paired with, or had a history of differential reinforcement with regard to, valenced images or stimuli that were themselves paired with such images (e.g., Pokémon 1 and 4).

Overall, the current work sheds light on a number of important issues related to the study of evaluation. As noted previously, a majority of EC research has sought to alter the functions of a single stimulus or set of stimuli via directly trained contingencies (see Hofmann et al., 2010). Our findings are striking insofar as they reveal that evaluative outcomes on both direct and indirect measures can be obtained even when no such contingencies are formed. To illustrate this more clearly, consider participants who failed the derivation test. In each case participants correctly identified the meaning of the two contextual cues and related *Pokémon1-Same-Pokémon2*, *Pokémon2-Same-Pokémon3*,

Pokémon4-Same-Pokémon5, *Pokémon5-Same-Pokémon6*, as well as *Pokémon1-Opposite-Negative* and *Pokémon4-Opposite-Positive*. Put another way, they demonstrated evidence of having generated a set of directly trained relations and responded to these relations with 100% accuracy across repeated blocks of trials. Nevertheless, and despite this history of learning, the fail group did not self-report any preferences for the various Pokémon characters. Moreover, they responded in a largely incoherent fashion on both the IRAP and IAT. The absence of evaluative responding for this group cannot reflect a simple failure to pay attention during training, engage with the task, understand what was required of them, or form stimulus relations based on contextual cues. Indeed, if any of these factors were responsible then it is unlikely that the exceptionally high performances that emerged and persisted throughout training would have occurred at all.

Rather it appears that participants who passed the derivation test engaged in an additional number of untrained behaviours relative to their counterparts who failed that task. Not only did they generate the above set of relations but also spontaneously formed a series of novel relations between those stimuli in a manner that was never explicitly trained. According to RFT, this capacity to weave derived relations between stimuli represents a form of generalized operant behaviour learned early on in our development (Barnes-Holmes & Barnes-Holmes, 2000). When approached from this perspective, it appears that participants arrived at the laboratory with a rich and extended history of arbitrarily applicable relational responding that generalized to novel, arbitrary stimuli (such as the symbols that served as contextual cues and with which the Pokémon characters and valenced images were related). If this assumption is correct, and relational responding is the basic functional unit underlying (implicit) cognition as well as other complex human behaviours, then the current data raise a potentially important question: why did such a large number of participants fail to show evidence for derivation when tested?

In explaining this outcome it may be important to pay particular attention to the methodology that we used to train relations between the various Pokémon characters. A common strategy in the RFT literature is to provide participants with multiple blocks of training followed by a derivation test in which mutual and combinatorial entailment is assessed. For example, Pokémon 3 may be presented with Pokémon 1 and participants asked to indicate whether those stimuli are related using the ‘Same’ or ‘Opposite’ cue. If participants fail to achieve some pre-specified response criteria during testing they are typically re-exposed to another round of training and testing until either evidence of derivation is obtained or the maximum number of trials exceeded. One potential drawback of this approach - at least in the context of the current work - is that the repeated co-occurrence of mutual and combinatorially entailed stimuli across successive test blocks allows for a purely associative explanation of the above results. In other words, directly pairing stimuli during the test phase introduces the possibility that derivation is nothing more than a second-order respondent process (see Dymond & Rehfeldt, 2000). Equally, a participant who is repeatedly exposed to a derivation test until the predicted response “emerges” may recognize that their performance is unsatisfactory (e.g., “*I keep encountering the same task; I must be responding incorrectly*”). When these repeated exposures to the derivation test are combined with the procedural limitations of the MTS protocol (Dymond & Rehfeldt, 2000) it becomes questionable whether any resulting transformation performances could be considered as genuinely derived.

In order to eliminate this possible criticism, participants in the current study immediately progressed from the derivation test to the various measures of evaluation - regardless of whether they passed or failed that task. Moreover, half of the participants moved straight from relational training to the evaluative measures without exposure to a derivation test at all. Although this strategy removes a potentially confounding source of

stimulus co-occurrence, it appears to come at a cost. Numerous studies have provided evidence for the “delayed emergence” of derived relations (e.g., Sidman, 1994; Holth & Arntzen, 1998; Wang, McHugh & Whelan, 2012) while others have shown that repeated practice across multiple training and testing blocks increases the number of participants who eventually derive (Barnes & Keenan, 1993; Devany, Hayes, & Nelson, 1986). As such, it is possible that the linear nature of the current work may have inflated the number of participants in the fail group given that we did not re-cycle individuals back through training if they failed the derivation test. Indeed, it may be the case that a significant number of the fail group could have shown evidence of derivation if repeated training and testing was provided. Future work could explore this possibility by systematically manipulating the amount of additional training participants receive if they fail to derive in the first instance.

On balance, it could be argued that Pokémon 3 and 6 were presented with valenced images during the derivation test and perhaps it was these pairings that were responsible for the aforementioned effects. If this assumption is correct then only participants who encounter the derivation test *before* the IAT and self-report procedures should have shown the predicted outcomes. Yet we found that completing the derivation test before or after the various measures of evaluation did not influence performance on those tasks. Thus it appears that a formal test for derivation is not necessary in order for a transformation of functions to occur. It is also worth noting that Pokémon 3 and 6 were presented an equal number of times with positive and negative images during the test and no corrective feedback was provided. As such, participants should have responded ambivalently towards both stimuli from an associative perspective. Once again this was not the case. Rather it seems that directly training a set of stimulus relations is insufficient to produce the observed effects - participants must also show evidence for the formation of derived relations as well (for related findings see Barnes-Holmes et al., 2004; O’Toole et al., 2007).

In addition to the above conceptual issues, our data also illustrate the advantages afforded by the IRAP in the measurement of laboratory induced histories of learning. Until now, it has not been possible to identify the specific stimulus relations that contributed to indirect effects due to the relativistic nature of the IATs used in Experiments 1 and 2. Consider, for example, the finding that participants responded with greater speed when they had to categorize Pokémon 3 with positive words and Pokémon 6 with negative words than vice versa. In this case, the resulting IAT effect could be interpreted in one of three different ways. Participants may have (a) liked Pokémon 3 and disliked Pokémon 6 (b) liked both Pokémon 3 and 6, but the former more so than the latter, or (c) disliked Pokémon 3 and 6, but the latter more so than the former. In other words, the IAT can indicate that X is preferred to Y, but it cannot discriminate the degree to which X and Y are individually liked or disliked. Consequently, while the task may be relatively simple to administer and be characterized by a high degree of psychometric reliability and predictive validity (Greenwald, Poehlman, Uhlmann & Banaji, 2009) these benefits appear to come at a cost (i.e., an inability to specify the precise relationship between the stimuli under investigation).

In order to circumvent this methodological limitation a second generation of IAT variants has recently evolved within the research literature. Examples include the Go/No-Go Association Task (Nosek & Banaji, 2001), Single Category IAT (Karpinski & Steinman, 2006), and Sorting Paired Features Task (Bar-Anan, Nosek & Vianello, 2009). Critically, however, and despite their ability to capture responses in a non-relativistic fashion, these measures appear to focus primarily on stimulus relations at a single level of complexity (i.e., coordination). For example, none of methodologies listed above can directly target responses that involve opposition, comparison, causality or hierarchy, nor do they ask participants to confirm or deny, in a relatively direct way, the relation under investigation. Thus an alternative methodology is needed that can provide a more precise measure of brief and

immediate relational responding than currently available elsewhere. The IRAP appears to be one such method.

To our knowledge the IRAP stands alone as the only indirect procedure that is capable of capturing different types of stimulus relations at differing levels of complexity in a non-relativistic fashion. In doing so, this task may permit a more fine-grained assessment of the relations that drive indirect effects than that afforded by the IAT or comparable measures (see Hughes & Barnes-Holmes, in press). With respect to the current work, for example, the IRAP revealed that participants responded quickly and accurately when they had to relate *Pokémon3-Same-Positive* and *Pokémon6-Same-Negative* but not when they had to relate *Pokémon3-Opposite-Negative* or *Pokémon6-Opposite-Positive*. These effects not only control for multiple interpretations introduced by the IAT but also provide further evidence for the transformation of function through stimulus relations. Indeed, it appears that indirect task performance across both tasks was governed by a combination of positive evaluative responding towards Pokémon 3 and negative evaluative responding towards Pokémon 6.

Interestingly, however, these effects seem to be driven primarily by affirming the derived relations rather than their negation. Several potential explanations present themselves. First, and as outlined in Chapter 2, participants may simply have a longer history of relating stimuli on the basis of sameness, similarity, or identity relative to other types of relations. Thus in contexts where a response has to be emitted with relative speed and precision (e.g., IRAP test blocks) coordination relations may – all things being equal – be at higher response strength or probability than those that involve opposition. Second, performance on the IRAP could have reflected the participants' history of derivation with respect to the Pokémon characters themselves. For instance, while Pokémon 1 was directly trained as the *Opposite-Negative* and Pokémon 4 as the *Opposite-Positive* participants may have derived based on coordination (e.g., “*Pokémon 3 is positive*” and “*Pokémon 6 is*

negative”) rather than opposition (e.g., “*Pokémon 3 is not bad*” and “*Pokémon 6 is not good*”). If correct, then derived coordination relations may have been at a higher response strength than those involving opposition, and in turn, been emitted with greater speed and precision on the task.

Third, laboratory induced histories of learning generated within a single session of training may give rise to relatively “weaker” indirect effects than their pre-existing counterparts based on race, gender, sexuality, consumer preferences or political orientation. For instance, and as noted in Chapter 2, newly established histories may be more sensitive to the alternating response contingencies encountered on indirect procedures such as the IAT and IRAP. Whereas the IAT is characterized by a single oscillation between consistent-inconsistent blocks, the architecture of the IRAP subjects these newly formed histories to at least four repeated reversals between competing relational contingencies. In doing so, the IRAP may function to undermine “weakly” established (opposition) relations to a greater degree than those at relatively higher response strengths (coordination) (see Hughes & Barnes-Holmes, 2011 for related findings).

To summarise, our work indicates that an experimental analysis framed solely in terms of simple stimulus pairings may not be enough – at least where verbally sophisticated humans are concerned. Unlike their non-human counterparts, people may spontaneously weave complex webs of derived relations that connect stimuli to one another in a variety of novel but predictable ways. Perhaps most importantly, these derived relations can allow stimuli to acquire new psychological functions in the absence of any direct history of learning. Consequently, a more complete understanding of how humans come to like and dislike stimuli may require that derived stimulus relating be elevated to centre stage within this research area.

Chapter 4: A Derived Transformation of Functions through Coordination and Opposition Relations as Measured by the IAT, Affective Priming and Self-Report Tasks

4.1 Experiment 4

Although the foregoing findings are consistent with our relational account of evaluative responding, one could still object that these claims only apply to a single set of conditioned (Pokémon) and unconditioned stimuli (valenced images) and may not generalize to other events in the environment. To control for this possibility, the current study examined whether evaluative outcomes would also be obtained when a new indirect measure (affective priming) and set of CSs (brand names) and USs (valenced adjectives) were employed. Once again, contextual cues meaning ‘Same’ and ‘Opposite’ were generated and used to form two coordination relations - this time comprised of three fictitious brand names (*Pardal-Same-Zatte-Same-Ettalas*; *Ciney-Same-Witkap-Same-Gageleer*). Thereafter, and for half of the participants, the first stimulus in either relation was related to valenced words using the ‘Same’ cue (*Pardal-Same-Positive* and *Ciney-Same-Negative*) while the other half were exposed to an opposition relation between those same stimuli (*Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*).

In adopting this design we sought to exert fine-grained contextual control over evaluative responding. If the relational networks are formed as predicted and a transformation of functions occurs, then two distinct patterns of behaviour should emerge. On the one hand, generating a coordination relation between *Pardal-Positive* and *Ciney-Negative* should result in both stimuli being evaluated in a direction specified by the valenced images they were related to (i.e., *Pardal* should be rated positively and *Ciney* negatively). Moreover, the second and third stimuli in the either relation should also elicit evaluative responses in accordance with the derived relation they participate in (e.g., *Zatte* and *Ettalas* should be “liked” whereas *Witkap* and *Gageleer* “disliked”). On the other hand, this pattern of responding should be

completely reversed when Pardal is related as *Opposite-Positive* and Ciney as *Opposite-Negative* (i.e., Pardal, Zatte and Ettalas should all be rated negatively while Ciney, Witkap and Gageleer rated positively). By systematically manipulating how participants respond to the same stimuli we aim to validate our previous findings and show that a transformation of functions is not unique to certain items or measures but rather reflects the operation of a general and genuine learning process.

It is worth noting that a majority of previous RFT studies have focused on “proof-of-principle” effects within the laboratory in order to demonstrate that functions may be transformed through contextually controlled relations (for a review see Dymond & Rehfeldt, 2000). While an important first step, the next generation of such studies will need to demonstrate that this behavioural process also predicts (and ultimately influences) “real-world” outcomes. One potential test-bed for this work may lie in the domain of consumer science. A common theme in this research area (and advertising, event sponsorship, and product placement more generally) is that the direct pairing of a commercial product with affective stimuli will increase the probability of that product being purchased in the future (see De Houwer, 2009b; Gibson, 2008; Sweldens, Van Osselaer & Janiszewski, 2010). A more interesting question - at least in the context of the current research - is whether people will also purchase a product simply because it participates in a derived relation with emotionally laden stimuli. Preliminary research in this vein seems to support this conclusion. For example, Barnes-Holmes and colleagues (2000) generated two equivalence classes consisting of a valenced adjective, non-sense syllable and novel brand name (e.g., *Cancer-Vek-Brand X* and *Holidays-Zid-Brand Y*). Thereafter, college students were asked to taste two identical soft drinks - one labelled Brand X and the other Brand Y - and rate their respective pleasantness. The authors found that subjects who passed an equivalence test subsequently rated Brand Y more positively than Brand X. In a follow-up study, Barnes-Holmes and

Smeets (2003) exposed Dutch schoolchildren to a similar protocol. Participants who passed an equivalence test not only rated Brand Y as more positive than its counterpart but also opted to taste that soft-drink first. When taken together, these findings suggest that stimulus equivalence procedures may not only allow for a transfer of function from a CS1 (socially established emotive word) via a CS2 (nonsense syllable) to a CS3 (brand name) but also influence the selection of consumer products. In Experiment 4 we sought to extend this work and determine whether similar findings would also emerge when increasingly complex relations are involved.

Towards this end, participants were informed that they would take part in a consumer research study involving a number of European brand products that had recently arrived on the market. They were then exposed to contextual cue and relational training, as well as the various measures of evaluation, after which, they were told that the study was over. Prior to their departure, however, six identical containers were presented - each labelled with one of the fictitious brand names previously encountered during the task. Participants were offered the opportunity to choose any three samples of the products as a “thank you” for their time and effort. We predicted that brand selection would reflect an interaction between (a) the type of relation established between Pardal/Ciney and valenced words as well as (b) performance on the derivation test. Put simply, participants should take the samples labelled Pardal, Zatte and Ettalas when they were exposed to *Pardal-Same-Positive* and *Ciney-Same-Negative* training. However, Ciney, Witkap and Gageleer should be the preferred products when *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative* relations are trained. Critically, these patterns of responding should only emerge for participants who successfully pass the derivation test. Indeed, their counterparts that fail to derive should select randomly from the six available options.

4.2 Method

Participants and Design

One hundred and fourteen undergraduates (67 female), ranging in age from 18 to 44 years ($M = 20$, $SD = 4.2$) participated as a partial fulfilment of course requirements and were each tested individually (). A 2(*Relation*: coordination vs. opposition) x 2(*Measure*: IAT vs. Affective Priming) x 2(*Training*: one vs. two sessions) design was employed, with all three variables manipulated between-participants. Two additional method factors were also counterbalanced; task order and IAT block order. Data from eleven individuals who failed to meet the training criteria as well as eight participants who demonstrated excessively high error rates on the affective priming task (> 50%) were excluded from analyses.

Materials

Stimuli. Prior to the study, twenty (non-participating) undergraduates were presented with a set of eighteen nonsense words and asked to provide an evaluative rating for each using a scale ranging from -3 (Negative Feelings) to +3 (Positive Feelings). The six names (*Pardal*, *Zatte*, *Ettalas*, *Ciney*, *Witkap*, and *Gageleer*) deemed the most neutral were selected as CSs (*Mean ratings* = .42, .47, .22, .06, .32 and .1 respectively) while the USs consisted of five positive (*delicious*, *fresh*, *tasty*, *sweet*, *yummy*) and five negative adjectives (*disgusting*, *stale*, *nasty*, *sick*, *rotten*)⁹.

Indirect procedures. For the IAT, the two brand names “Gageleer” and “Ettalas” served as one set of target stimuli and the words “Good” and “Bad” as another. Six positively and negatively valenced adjectives served as one set of attribute stimuli (*nice*, *tasty*, *delicious*, *sweet*, *fresh*, *yummy* versus *sick*, *horrible*, *disgusting*, *nasty*, *stale*, *rotten*) and the words “Gageleer” and “Ettalas” as a second. For the affective priming task, we used four of the

⁹ Note that the above CSs were in fact the names of six Belgian beers. Critically, pre-testing indicated that none of the participants were familiar with any of these names prior to the study.

brand names as prime stimuli (*Ettalas, Zatte, Gageleer and Witkap*) and four positive (*delicious, tasty, yummy, nice*) and negative words (*disgusting, rotten, sick, vomit*) as target stimuli.

Procedure

Participants were informed that they would take part in a study concerning brand products that had just been introduced to the European marketplace. They were also told that it was unlikely that they had ever seen any of these brands before, but that they would be provided with an opportunity to learn about them during the task. They then completed five different experimental phases in the following order; contextual cue training, relational training, a derivation test, (in)direct procedures and a behavioural choice task.

Contextual Cue Training

In the first part of the experiment, the relational functions of ‘Same’ and ‘Opposite’ were established for two arbitrary symbols using a broadly similar procedure as before.

Relational Training

We then set out to generate two coordination relations, each comprised of three fictitious brand names (*Pardal-Same-Zatte-Same-Ettalas and Ciney-Same-Witkap-Same-Gageleer*) using a procedure that was similar in many respects to that employed previously (see Figure 4.1). Training consisted of four separate phases, each comprised of a minimum of one and a maximum of three blocks of 50 trials. On-screen instructions informed participants that they would be presented with two brand names on the upper left and right sides of the screen, as well as the two contextual cues they previously encountered at the bottom of the screen. Their task was to “*determine the relationship between the two brand products at the top of the screen using the symbols you have just learned about*”.

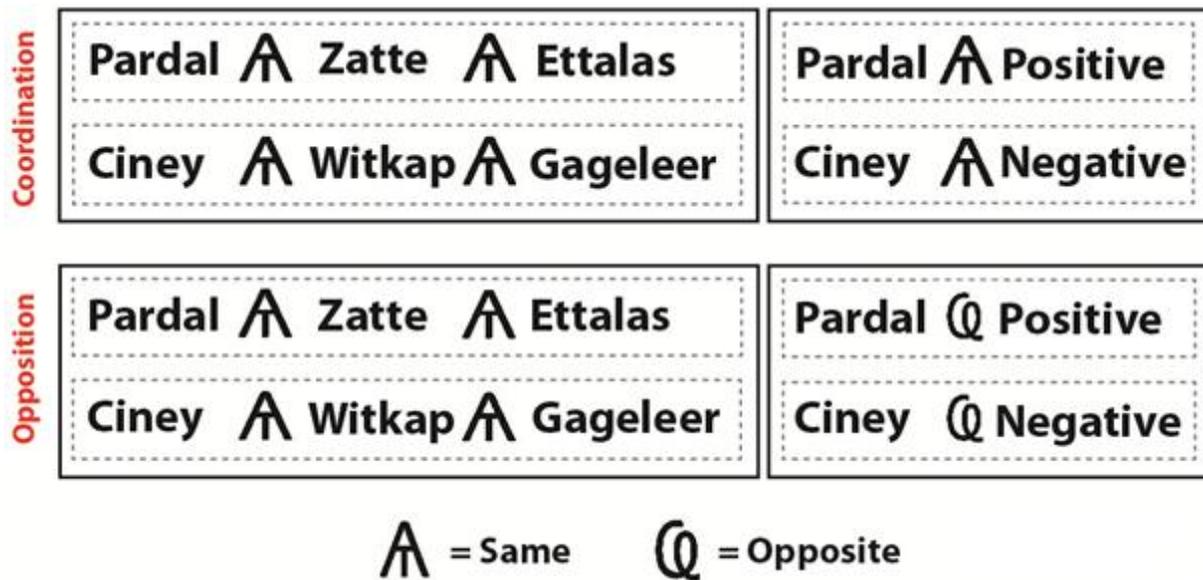


Figure 4.1. Schematic representation of the trained relations in Experiment 4. Two, three member coordination relations were established each comprised of three fictitious brand products (*Pardal-Same-Zatte-Same-Ettalas* and *Ciney-Same-Witkap-Same-Gageleer*). Half of the participants were then trained to relate *Pardal-Same-Positive* and *Ciney-Same-Negative* while the other half were trained to relate *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*.

During the first phase of training, four relations were established by differentially reinforcing the selection of one of the two contextual cues in the presence of a specific stimulus combination (*Ciney-Same-Witkap*; *Pardal-Same-Zatte*; *Witkap-Opposite-Pardal*; *Zatte-Opposite-Ciney*). For example, when Ciney and Witkap or Pardal and Zatte were displayed together, selecting the ‘Same’ cue caused “Correct” to appear for 1000ms, followed by an inter-trial interval (ITI) and the next trial. Making an incorrect response - such as selecting ‘Opposite’ in the presence of the above stimuli - caused error feedback to appear in the middle of the screen. In order to progress to the following trial participants were required to select the correct (‘Same’) cue. In contrast, when Witkap and Pardal or Ciney and Zatte were presented together, selecting the ‘Opposite’ cue was reinforced and the ‘Same’ cue punished. Within a block of trials the allocation of the two symbols to the lower left and right sides of the screen, as well as presentation of a stimulus combination was varied in a quasi-random order (see Figure 4.2).

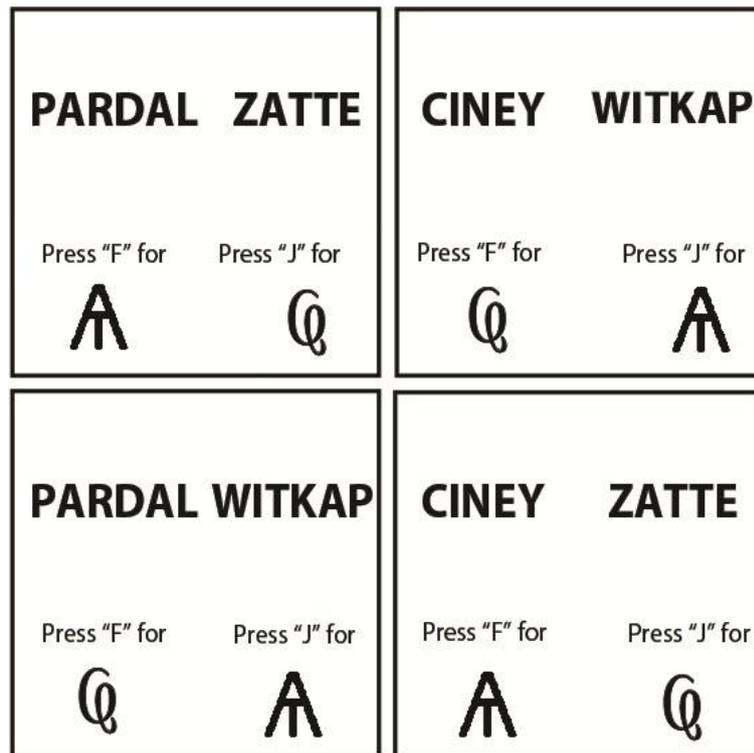


Figure 4.2. Examples of the four trials involved in the first phase of relational training. Each trial displayed two brand names at the top of the screen and the two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being presented in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).

The second phase of training was identical to the first with the exception that four additional relations were generated (*Witkap-Same-Gageleer*; *Zatte-Same-Ettalas*; *Witkap-Opposite-Ettalas*; *Zatte-Opposite-Gageleer*). In the third phase participants were re-exposed to all eight relations trained during phases one and two. Thereafter, and in the final phase, an evaluative function was established for the first brand name in both relations. For half of the participants, this function was established by generating a coordination relation between a brand name and valenced words. This was achieved by presenting either Pardal and a positive word or Ciney and a negative word at the top of the screen along with the two contextual cues at the bottom of the screen. Selecting the ‘Same’ cue resulted in “Correct” being displayed (followed by the ITI and the next trial) while selecting the ‘Opposite’ cue caused error feedback to appear. For the other half of participants, evaluative functions for Pardal and Ciney were established via an opposition relation. That is, selecting the ‘Opposite’ cue when

presented with the above stimuli caused positive feedback to appear onscreen while selecting the ‘Same’ cue resulted in error feedback¹⁰.

To proceed from one training phase to another, participants were required to achieve a mastery criterion of 20 consecutively correct responses on a block of training trials followed by a minimum of 20 out of 24 correct responses on a block of test trials. Failure to do so resulted in re-exposure to the task, or following a total of three training and testing blocks, in the participant being thanked, debriefed and their data discarded (eleven participants were removed on this basis). Finally, to determine if the amount of training influenced evaluative responding, half of the participants received two sessions of contextual and relational training across two separate days while the other half received a single session of training.

Test for Derived Relational Responding

Following training, participants completed a derivation test in order to determine if the relational networks were formed as predicted (and the functions established for Pardal and Ciney were transformed through those networks to Ettalas and Gageleer). Each trial presented Ettalas or Gageleer in addition to a positive or negative word and the two contextual cues. On-screen instructions asked participants to “*click on the symbol that describes the relationship between the two pictures at the top of the screen*” (an option to indicate “*I do not know the relationship*” was also provided). Clicking on a contextual cue or

¹⁰ In the current and previous studies reported here evaluative functions were established by extending relational networks, to include images or words from natural language, via frames of coordination or opposition. It may be tempting to argue therefore that this manipulation involves only entailment processes and not a transformation of function per se. It is important to remember, however, that according to RFT entailment processes are a type of transformation of function if only in a limited sense. Furthermore, the term “transformation of function” tends to be used when appetitive, aversive or other functions of the stimuli in a network (above and beyond entailment relations) are changed as a result of some procedure. To illustrate, consider a three-member equivalence class (A-B-C) and a procedure that then pairs A with the delivery of electric shock. We could say that we have extended the network to a four-member class that now includes shock as a stimulus. Although this description certainly captures entailed relations among shock and the three stimuli it does not capture the other types of functions that would likely emerge. For example, participants may well show signs of fear and engage in avoidance responding whenever any of the three stimuli are presented. In such cases, the term “transformation of function” is typically used because the emergence of fear and avoidance functions extends beyond basic entailment relations. In the current thesis the transformation of functions is assessed using a range of measures, such as the IAT, IRAP, priming, and behavioural choice tasks. Or more informally, each of these tasks is designed to determine how the various stimuli are evaluated emotionally rather than how they are simply categorized.

“I don’t know” button cleared all the stimuli from the screen, followed by an inter-trial interval of 500ms and the next trial. Testing involved a block of twelve trials with Ettalas and Gageleer presented an equal number of times in the presence of positive and negative words. No feedback was provided for any response emitted during this task.

For participants trained to relate *Pardal-Same-Positive* and *Ciney-Same-Negative*, a correct response was defined as the selection of the ‘Same’ cue when given Ettalas and a positive word; ‘Opposite’ when given Ettalas and a negative word; ‘Same’ when presented with Gageleer and a negative word; and ‘Opposite’ given Gageleer and a positive word. For participants trained to relate *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*, the reverse pattern of responding across all trial-types was required. In either case, a minimum of 10 out of 12 correct responses was required to pass the test and those who did not meet this criterion were defined as having failed the test.

Indirect Procedures

IAT. Half of the participants were administered a similar IAT as before in order to investigate brief and immediate relational responding towards Ettalas relative to Gageleer. If the relational networks were established in-line with prior training then response latencies should be shorter during consistent relative to inconsistent trials. For participants trained to relate *Pardal-Same-Positive* and *Ciney-Same-Negative*, faster responding should be obtained when Ettalas has to be related with positive relative to negative stimuli and Gageleer with negative relative to positive stimuli. For participants trained to relate *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*, a reverse pattern of responding was expected (i.e., faster responding when Ettalas has to be related with negative relative to positive stimuli and Gageleer with positive relative to negative stimuli).

Affective priming. The affective priming task consisted of one practice and two test blocks. The second and third stimuli from both relations were deployed as primes and five

positive and negative words as targets. Participants were informed that one of the brand names they had previously encountered would appear in the middle of the screen, rapidly followed by a word. Their task was to indicate the valence of that word as quickly and accurately as possible, by pressing the “L” key if it was positive or the “S” key if it was negative. No further mention of the primes was made in the instructions. Each trial started with the presentation of a fixation cross for 500ms, which was replaced by a brand name for 200ms. Immediately thereafter, a target stimulus was presented, and remained on-screen until a response was emitted. The next trial started after an inter-trial interval of 1000ms.

The task began with six practice trials during which one of the brand names was randomly assigned as a prime and either a positive or negative word as a target. Following practice, and to increase adherence to the speed criterion, participants were reminded that it was of the utmost importance that they respond as quickly as possible. Thereafter two blocks of 32 test trials were administered, during which each of the four stimuli was randomly combined with four positive and four negative targets. The order of these trials was randomized within each block.

For participants trained to relate *Pardal-Same-Positive* and *Ciney-Same-Negative*, faster responding was expected on trials in which Ettalas appeared with positive relative to negative stimuli and Gageleer with negative relative to positive stimuli. For participants trained to relate *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*, a reverse pattern of responding was expected (i.e., faster responding when Ettalas appeared with negative relative to positive stimuli and Gageleer with positive relative to negative stimuli).

Direct Procedures

CS ratings. Self-reported ratings for each of the brand names were assessed using a series of Likert scales. Participants were asked to rate the likeability of the brand names using

a response scale that ranged from -4 (Negative Feelings) to +4 (Positive Feelings) with 0 (Neutral) at its midpoint.

Strategy questions. Similar to previous experiments, a number of additional questions were presented concerning demand compliance and the meaning they ascribed to the contextual cues.

Behavioural choice task. Following the direct and indirect measures, participants were informed that the experiment was complete. Prior to debriefing and their departure, however, they were presented with six small boxes that were identical in size, shape and colour, each with the name of one brand printed on the top (the respective locations of the boxes in relation to one another were varied randomly across participants). As a reward for their participation, they were offered the opportunity to select three “samples” of the brand products to take home with them. Once participants had made their choice they were fully debriefed, thanked and dismissed.

4.3 Results

Data Preparation

Contextual cue meaning. Of the current sample, ninety three participants (98%) reported the relational functions of the contextual cues in-line with experimental expectations. On the one hand, forty nine participants (53%) rated the ‘Same’ cue as meaning “same”; twenty eight (30%) rated it as meaning “similar” while the remaining sixteen (17%) used terms such as “alike”, or “related”. On the other hand, fifty participants (54%) rated the ‘Opposite’ cue as “opposite”; twenty nine (31%) rated it as meaning “different” while the remaining fourteen (15%) used a variety of terms such as “dissimilar”, “not same” or “not connected”. The data for two participants who reported incorrect relational functions were removed prior to analysis. Reanalyzing the data with these participants included did not change any of the statistical conclusions reported below.

Demand compliance. Five of the ninety five participants (5.2%) indicated that they intentionally responded to the brands in-line with the presumed expectations of the experimenter. Data for these participants were removed prior to analysis. Once again, reanalyzing the data with these individuals included did not influence any of the obtained effects.

Preliminary analyses. Counterbalancing the amount of training received, order of IAT blocks as well as direct and indirect tasks produced no significant effects. Consequently, analyses were collapsed across these factors. Nineteen participants (20%) failed the derivation test while seventy six (80%) passed the test.

Self-Reported Ratings

A series of mean likeability scores for the brand products in the first relation are presented in Figure 4.3 while the scores from the second relation are presented in Figure 4.4. Visual inspection of the graphs revealed that relational training was once again successful, with self-reported ratings varying according to whether a coordination or opposition relation was established between a brand name and valenced words. For instance, relating *Pardal-Same-Positive* and *Ciney-Same-Negative* resulted in Pardal, Zatte and Ettalas being rated positively and Ciney, Witkap and Gageleer negatively. However, when those same stimuli were related using the ‘Opposite’ cue evaluative responding was completely reversed – with Pardal, Zatte and Ettalas rated negatively and Ciney, Witkap and Gageleer rated positively.

This description of the data was supported by the results of a 3(*Brand; Pardal, Zatte, Ettalas*) x 2(*Relation; coordination vs. opposition*) x 2(*Derivation test performance; pass vs. fail*) mixed-model ANOVA. Analyses revealed a main effect for relation, $F(1, 83) = 125.6, p = .001, \eta^2_{Partial} = .6$, and brand, $F(2, 83) = 5.2, p = .006, \eta^2_{Partial} = .06$, a two-way interaction between brand and relation, $F(2, 83) = 14.8, p = .001, \eta^2_{Partial} = .15$, as well as a three-way interaction between brand, relation and test performance, $F(2, 83) = 6.7, p = .002, \eta^2_{Partial} =$

.08. In order to specify this three-way interaction, evaluative scores for participants who passed the derivation test were assessed separately from those who failed.

With respect to the pass group, a main effect emerged for brand, $F(2, 67) = 3.5, p = .03, \eta^2_{\text{Partial}} = .05$, with planned contrasts revealing that the second stimulus was rated more positively than the third, $F(1, 67) = 7.1, p = .009, \eta^2_{\text{Partial}} = .1$. There was also a significant main effect for relation, $F(1, 67) = 336.8, p = .001, \eta^2_{\text{Partial}} = .83$, with self-reported ratings varying as a function of the contextual cue used to relate the CS and US. In particular, participants responded positively towards Pardal when it was related to positive words using the ‘Same’ cue ($M = 3.3, SD = 1.5$), but negatively when it was related to positive words using the ‘Opposite’ cue ($M = -2.9, SD = 1.4$), $F(1, 68) = 299.3, p = .001, \eta^2_{\text{Partial}} = .82$. A similar set of findings emerged for Zatte, $F(1, 68) = 298.8, p = .001, \eta^2_{\text{Partial}} = .82$, and Ettalas, $F(1, 68) = 121.3, p = .001, \eta^2_{\text{Partial}} = .64$, such that participants responded positively to Zatte ($M = 3.0, SD = 1.2$) and Ettalas ($M = 2.3, SD = 2.2$) when *Pardal-Same-Positive* and negatively to Zatte ($M = -2.7, SD = 1.5$) and Ettalas ($M = -2.9, SD = 1.6$) when *Pardal-Opposite-Positive*. All evaluative scores were independently significant from zero (all $ps < .001$).

Interestingly, participants who failed the derivation test showed a main effect for relation, $F(1, 16) = 14.2, p = .002, \eta^2_{\text{Partial}} = .47$, as well as a two-way interaction between brand and relation, $F(2, 16) = 8.7, p = .001, \eta^2_{\text{Partial}} = .35$. Follow-up one-way ANOVAs indicated that they responded positively to Pardal when it was related as *Same-Positive* ($M = 3.2, SD = .75$) and negatively when it was *Opposite-Positive*, ($M = -1.1, SD = 1.4$), $F(1, 17) = 44.8, p = .001, \eta^2_{\text{Partial}} = .73$. Likewise, when *Pardal-Same-Positive* Zatte was also rated positively ($M = 1.8, SD = 1.4$) and when *Pardal-Opposite-Positive* that same stimulus was rated neutrally ($M = -.25, SD = 1.6$), $F(1, 17) = 7.1, p = .02, \eta^2_{\text{Partial}} = .31$. Critically, however, no evaluative scores emerged for the final stimulus (Ettalas) regardless of the type of relation

established between Pardal and positive words, $F(1, 17) = .11, p = .7$. Moreover, a series of one-sample t-tests indicated that only the evaluative ratings for Pardal ($p = .001$) and Zatte ($p = .03$) in the *Same-Positive* relation and Pardal ($p = .03$) in the *Opposite-Positive* relation were independently significant from zero.

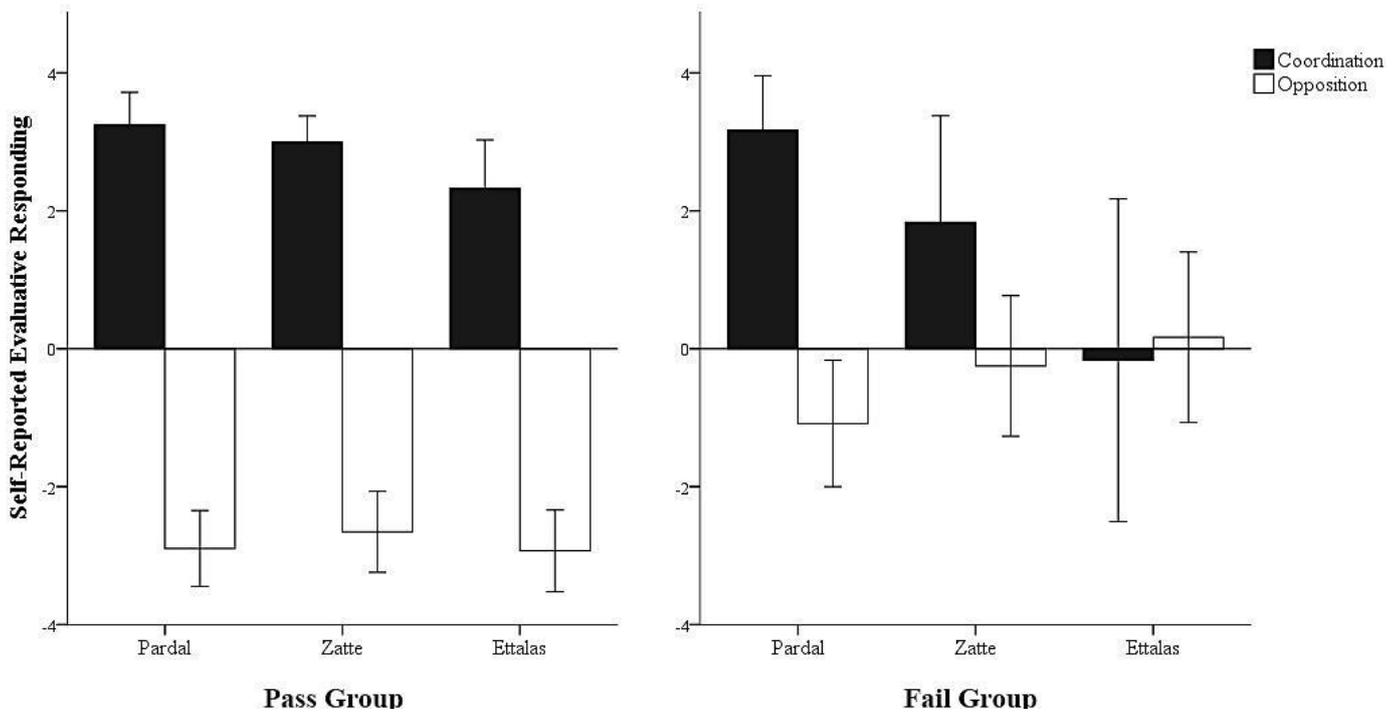


Figure 4.3. Mean likeability scores for Pardal, Zatte and Ettalas as a function of relation (coordination vs. opposition) and derivation test performance (pass vs. fail). Error bars indicate standard errors.

When the second stimulus relation (*Ciney-Witkap-Gageleer*) was submitted to the same analyses, a comparable set of findings emerged. As before, a main effect was obtained for relation, $F(1, 83) = 64.7, p = .001, \eta^2_{Partial} = .44$, and brand, $F(2, 83) = 4.2, p = .02, \eta^2_{Partial} = .05$, a two-way interaction between brand and relation, $F(1, 83) = 21.4, p = .001, \eta^2_{Partial} = .21$ and a three-way interaction between brand, relation and derivation test performance, $F(2, 83) = 7.2, p = .001, \eta^2_{Partial} = .10$. Once again, evaluative scores were analyzed separately based on whether participants passed or failed the derivation test. With respect to the pass group, a main effect emerged for relation, $F(1, 67) = 195.8, p = .001, \eta^2_{Partial} = .74$. Follow-up analyses indicated that participants responded negatively to Ciney when it was related as

Same-Negative ($M = -2.9$, $SD = 1.9$), but positively when it was related as *Opposite-Negative* ($M = 2.9$, $SD = 1.4$), $F(1, 68) = 189.6$, $p = .001$, $\eta^2_{\text{Partial}} = .74$. This was also true for Witkap, $F(1, 68) = 162.7$, $p = .001$, $\eta^2_{\text{Partial}} = .71$, and Gageleer, $F(1, 68) = 152.1$, $p = .001$, $\eta^2_{\text{Partial}} = .69$, with participants responding negatively to Witkap ($M = -2.7$, $SD = 1.8$) and Gageleer ($M = -2.8$, $SD = 1.9$) when *Ciney-Same-Negative* and positively to Witkap ($M = 2.9$, $SD = 1.6$) and Gageleer ($M = 2.8$, $SD = 1.8$) when *Ciney-Opposite-Negative*.

With respect to the fail group, a main effect also emerged for relation, $F(1, 16) = 5.0$, $p = .04$, $\eta^2_{\text{Partial}} = .24$, as well as a two-way interaction between brand and relation, $F(2, 16) = 4.2$, $p = .02$, $\eta^2_{\text{Partial}} = .21$. Follow-up one-way ANOVAs indicated that participants responded negatively to Ciney when it was related as *Same-Negative* ($M = -2.8$, $SD = 1.2$) and positively when it was *Opposite-Negative*, ($M = .67$, $SD = 2.3$), $F(1, 17) = 11.6$, $p = .004$, $\eta^2_{\text{Partial}} = .42$. Critically, however, no evaluative scores emerged for the second (Witkap) and third stimuli (Gageleer) regardless of the type of relation established between Ciney and negative words (all $ps > .5$). Indeed, a series of one-sample t-tests indicated that only the evaluative ratings for Ciney ($p = .002$) in the *Same-Positive* relation was significantly different from zero.

When taken together, these findings indicate that participants who passed the derivation test transformed the evaluative functions through the derived relations as anticipated. Although their counterparts in the fail group also showed evidence of evaluative responding, ratings were largely confined to stimuli directly related to valenced words (i.e., a transformation of functions through combinatorially entailed relations was absent when participants failed the derivation test).

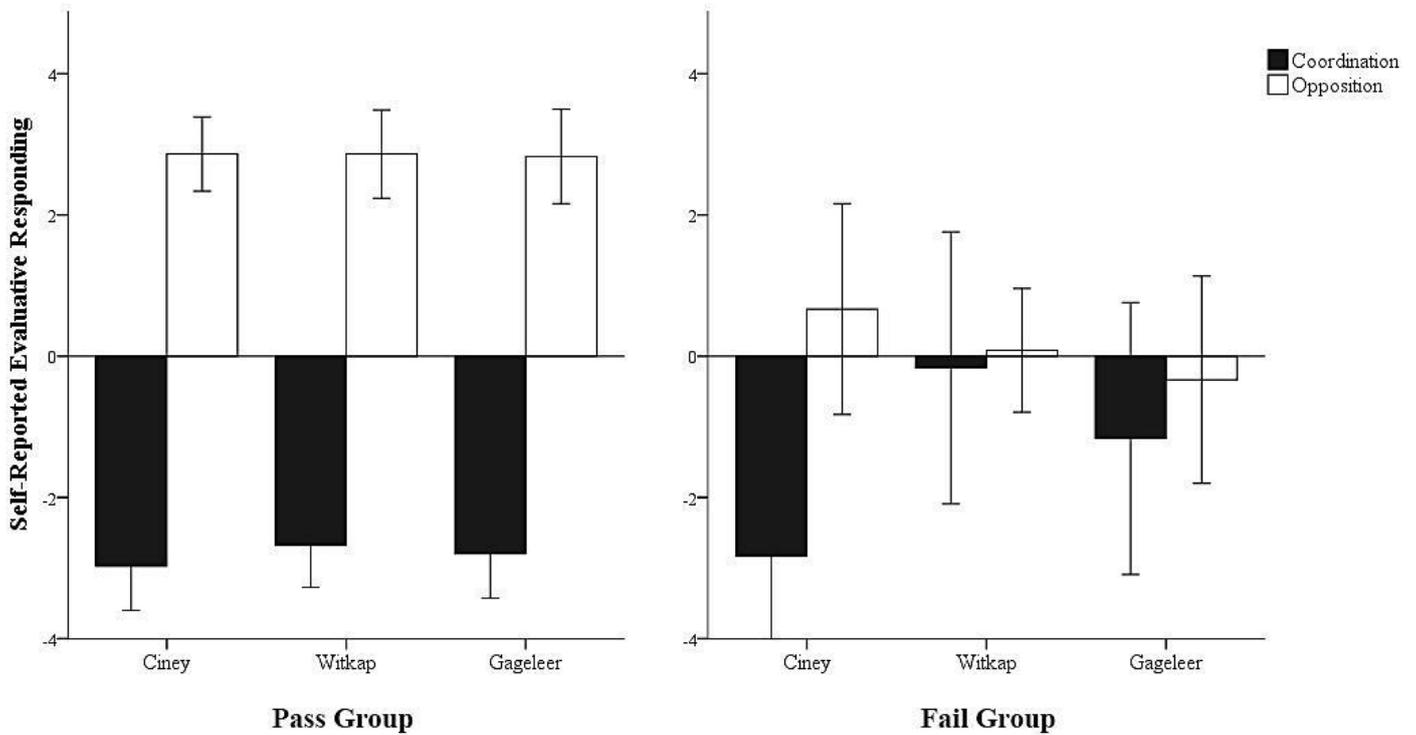


Figure 4.4. Mean likeability scores for Ciney, Witkap and Gageleer as a function of relation (coordination vs. opposition) and derivation test performance (pass vs. fail). Error bars indicate standard errors.

Indirect Procedures

IAT. Data from the IAT was prepared in a similar fashion as before. Scores were calculated so that positive values reflected a response bias favouring Ettalas relative to Gageleer while negative values indicated the reverse pattern of responding. To investigate whether the functions established for Pardal and Ciney were transformed through their respective relations to Ettalas and Gageleer, IAT scores were submitted to a $2(\text{Relation}) \times 2(\text{Derivation test performance})$ ANOVA. A significant main effect was observed for relation, $F(1, 40) = 15.5, p = .001, \eta^2_{\text{Partial}} = .3$, as well as a two-way interaction between relation and test performance, $F(1, 40) = 14.3, p = .001, \eta^2_{\text{Partial}} = .28$. To specify this two-way interaction, IAT scores were assessed separately for those who passed versus failed the derivation test. With respect to the pass group, a significant IAT effect emerged that was consistent with a transformation of function, $F(1, 32) = 69.5, p = .001, \eta^2_{\text{Partial}} = .69$. More precisely, when Pardal was *Same-Positive* and Ciney *Same-Negative* a response bias emerged for Ettalas relative to Gageleer ($M = .53, SE = .07$) $t(17) = 7.7, p = .001$. Yet when Pardal was *Opposite-*

Positive and *Ciney Opposite-Negative* this response bias was completely reversed, such that the IAT effect favoured Gageleer relative to Ettalas, ($M = -.31, SE = .07$) $t(14) = -4.3, p = .001$. Consistent with our predictions, this pattern of responding was entirely absent when participants failed the derivation test (all $ps > .3$) (see Figure 4.5).

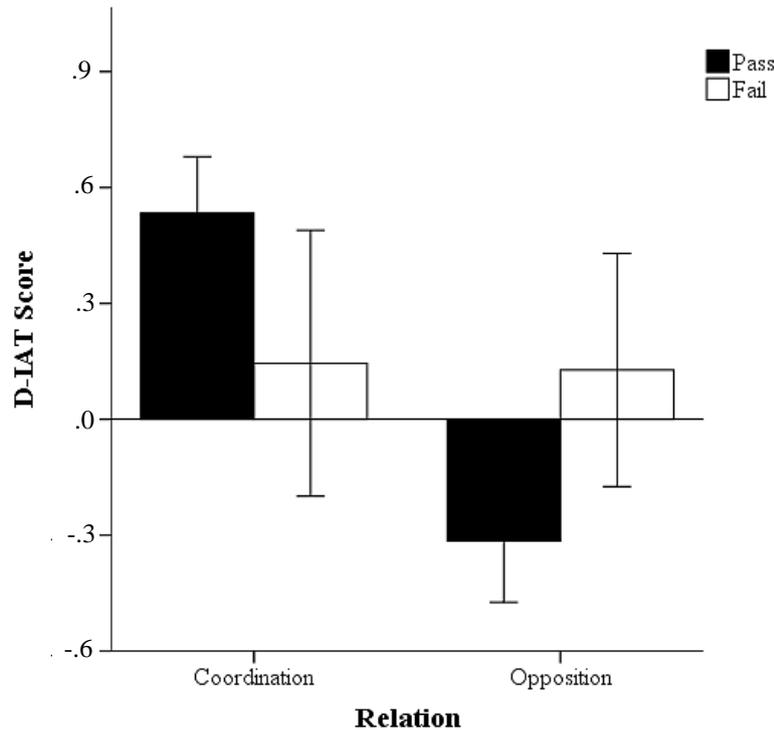


Figure 4.5. Mean D-IAT scores as a function of derivation test performance (pass vs. fail) and the type of relation established (coordination vs. opposition). Error bars represent standard errors.

Affective priming. Priming data was corrected for outliers by excluding trials on which the target was incorrectly classified (3.8%) or with reaction times below 300ms or greater than 1000ms (3.4%). Two evaluative scores were calculated, one for Ettalas and another for Gageleer, by subtracting the mean response latencies for trials in which a brand name appeared with negative words from those in which it appeared with positive words. In either case, a positive value reflects a response bias favouring that stimulus. Note that only the differences between these two scores - and not their absolute value - should be interpreted (e.g., a value of zero does not reflect a neutral evaluation). Given the restricted size of the fail

group, meaningful analysis across the various experimental conditions was not possible for priming data. Consequently, only data from the pass group is analyzed below.

When submitted to a $2(\text{Relation}) \times 2(\text{Prime})$ repeated measures ANOVA, a significant two-way interaction emerged between prime and relation, $F(1, 30) = 6.5, p = .02, \eta^2_{\text{Partial}} = .18$, indicating that performance on the task differed as a function of the relation established between the stimuli. On the one hand, a marginally more positive score emerged for Ettalas ($M = 31.5, SE = 12.3$) relative to Gageleer ($M = -.26, SE = .19$) when *Pardal-Same-Positive* and *Ciney-Same-Negative* – however this trend was non-significant, $t(19) = 1.6, p = .1$. On the other hand, a reverse pattern of responding emerged when *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*, with Gageleer ($M = 39.7, SE = 15.1$) yielding a significantly more positive evaluative score relative to Ettalas ($M = -5.5, SE = 13.5$), $t(14) = -2.19, p = .05$ (see Figure 4.6).

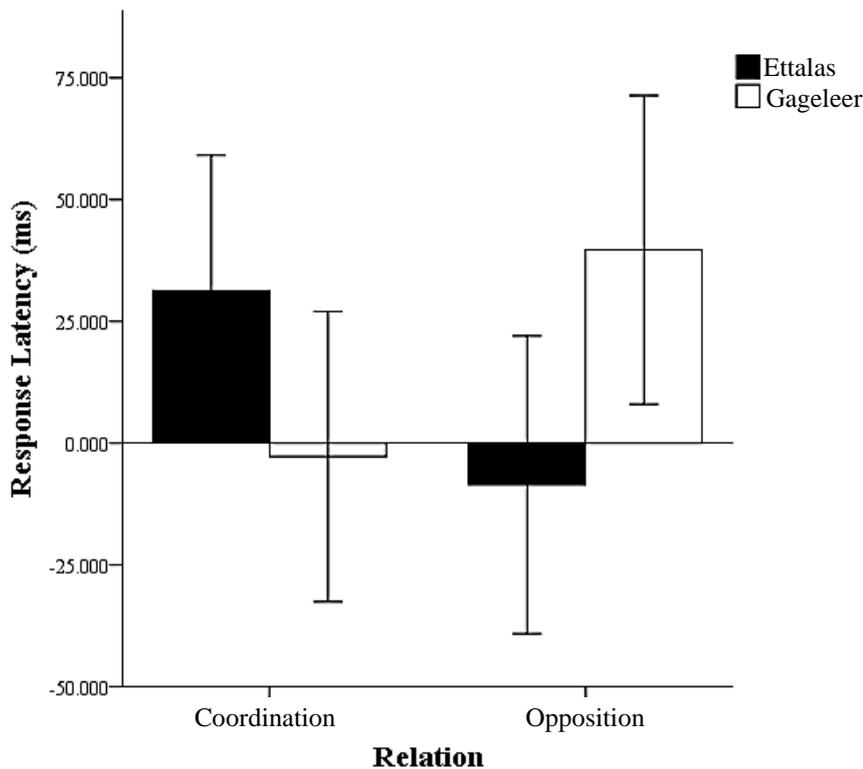


Figure 4.6. Mean response latencies for the positive and negative primes as a function of the type of relation established (coordination vs. opposition). Error bars represent standard errors.

Behavioural Choice Task

A correct response on the behavioural choice task was defined as the selection of Pardal, Zatte and Ettalas when participants were exposed to *Pardal-Same-Positive* and *Ciney-Same-Negative* training. In contrast, the selection of Ciney, Witkap and Gageleer was defined as a correct response when participants encountered *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative* training. In either case, the selection of any combination of brand products from across the two relational networks was defined as an incorrect response. 86% of participants who passed the derivation test selected the three correct brands in-line with the relations established among the stimuli during training. In contrast, 43% of those who failed the derivation test selected the three correct brands while 57% selected randomly from the six available options. When a Chi Square test was performed to examine whether brand selection was distributed differently across those who passed versus failed the derivation test a significant effect was obtained, $\chi^2(1) = 12.9, p = .001$. This indicates, based on the odds ratio, that the likelihood of selecting the three correct brands was 8.7 times higher if participants passed the derivation test than if they failed to do so.

4.4 Discussion

The above results further expand the remit of our relational account and demonstrate that derived stimulus relating influences how humans evaluate stimuli in the environment. Three core hypotheses received support in Experiment 4. The first was that self-reported ratings of the brands would be contingent upon two factors: the cue governing the trained relations as well as performance on a derivation test. Evidence for this first assumption can be found in the distinct pattern of outcomes noted above. For instance, when the selection of the ‘Same’ cue was reinforced in the presence of Pardal and positive words (or Ciney and negative words) participants responded positively to Pardal and negatively to Ciney. In addition, these evaluative functions were transformed through the respective relations such

that Zatte and Ettalas were also rated positively while Witkap and Gageleer rated negatively. However, a dramatically different picture emerged when the selection of the ‘Opposite’ cue was reinforced in the presence of those same stimuli. Participants now rated Pardal, Zatte and Ettalas negatively and Ciney, Witkap and Gageleer positively.

The second hypothesis was that similar findings would also be observed when two different indirect measures were administered (IAT and Affective Priming). This hypothesis was derived from prior work suggesting that histories of relational learning can be emitted with relative speed and accuracy (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010) and that these responses will cohere with other relational responses (such as those emitted on direct measurement procedures) when participants are not motivated to self-present or modify their behaviour to concord with experimental or social expectations (Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009; Barnes-Holmes, Murtagh, Barnes-Holmes & Stewart, 2010). Consistent with this prediction, participants who learned *Pardal-Same-Positive* and *Ciney-Same-Negative* subsequently produced an IAT effect favouring Ettalas over Gageleer. An entirely opposite response bias was engineered so that participants favoured Gageleer over Ettalas by simply training Pardal as *Opposite-Positive* and Ciney as *Opposite-Negative*. It is also worth noting that a priming effect was obtained for Ettalas relative to Gageleer when participants received coordination training (although this difference did not reach significance), and Gageleer relative to Ettalas when opposition training was provided. Remarkably, these effects emerged despite the fact that neither Ettalas nor Gageleer was ever paired with, nor had a history of reinforcement with regard to, valenced words or even stimuli that were themselves paired with such words (e.g., Pardal or Ciney).

The third and final hypothesis concerned the probability of selecting a brand product as a function of the derived relation it participated in. In-line with our predictions, a majority of participants chose to take Pardal, Zatte and Ettalas home with them following coordination

training and Ciney, Witkap and Gageleer following opposition training. It is important to keep in mind, however, that IAT, priming and behavioural choice performances were only evident when participants derived the untrained relations between the various brand products. Indeed, while the fail group liked and disliked stimuli that were directly related to valenced words they did not show any evidence of evaluative responding towards the combinatorially entailed stimuli. In addition, they showed no IAT effects and selected from the brands at near chance levels during the behavioural choice task. This is despite the fact that they passed repeated blocks of contextual cue and relational training with 100% accuracy. When taken together, the current work indicates that the extent to which people like or dislike stimuli depends on an interaction between their learning history and contextual factors. A transformation of function may not only explain how stimuli indirectly acquire evaluative functions but also why people may opt for one stimulus over another.

The above findings hold several interesting implications for marketing researchers and consumer psychologists alike. A persistent question in these respective areas is whether pairing a stimulus with positive or negative scenarios, events, celebrities or consequences will increase the probability of certain behaviours, such as buying a product or eating healthy instead of unhealthy foods. Much of this work has found that preferences for novel brands can be generated (Stuart, Shimp & Engle, 1987; Walther & Grigoriadis, 2004) or existing brands altered (Gibson, 2008) through stimulus pairings procedures such as evaluative conditioning. Furthermore, these preparations can also lead to important outcomes such as reduced alcohol (Houben, Schoenmakers & Wiers, 2010), or unhealthy food intake (Hollands et al., 2011). Despite this growing appreciation for the role of direct contingencies in shaping consumer choice behaviours, the impact of derived relational responding has yet to be explored with equal vigour. Indeed, to our knowledge, only two studies have tested, and demonstrated, that people will select a brand product simply because it participates in a

derived relation with other emotionally valenced stimuli (Barnes-Holmes et al., 2000; Smeets & Barnes-Holmes, 2003). In both cases, however, the authors focused exclusively on equivalence relations and measured responding using self-report and behavioural choice tasks. Our findings not only replicate these early studies but extend their scope by highlighting that a transformation of function can influence behaviour when more complex (opposition) relations and indirect procedures are involved.

Nevertheless, it should be noted that these “proof-of-principle” effects engineered in the laboratory do not automatically imply that people behave in the same way in their everyday lives. Rather, additional translational research is needed in order to extrapolate these findings from the laboratory to wider world, and in particular, demonstrate that the presence versus absence of derivation meaningfully discriminates between different groups and behavioural phenomena in ecologically valid settings (see also Critchfield, 2011). While social and consumer psychologists have adopted translational research as another tool in their intellectual arsenal, functional researchers have mainly focused on transforming avoidance functions (Dymond et al., et al., 2008), false memories (Guinther & Dougher, 2010) and respondent functions (Smyth et al., 2006), as well as sexual (Roche & Dymond, 2008), extinction (Roche et al., 2008) and discriminative functions (Dougher et al., 2007) through experimentally established relations.

While these early studies are unquestionably necessary, the next generation of work will need to extend laboratory findings to more naturalistic environments. For instance, it remains to be seen whether these evaluative effects also occur under conditions that more accurately reflect those encountered in the “real-world” (e.g., brief advertisements on the television, radio or internet; exposure to “samples” of a product followed by information that the products is *better than* brand X and *cheaper than* Brand Y). Likewise, will a transformation of function established in the lab influence consumer choice behaviours long

after and in settings that differ radically from those in which training originally took place (e.g., a supermarket)? While our contextual cue and relational training protocols may not represent a viable means to establish relational histories within naturalistic contexts, a modified version of Leader and colleagues (2000) respondent-type training task might offer a pragmatic alternative. Similar to commercials, billboards, and product placements, this task simply presented pairs of stimuli together in close spatial and temporal proximity ($A1 \rightarrow B1$; $A2 \rightarrow B2$; $B1 \rightarrow C1$ and $B2 \rightarrow C2$). When subsequently tested participants consistently show evidence of having formed derived relations between those stimuli (Smeets et al., 1997; Leader et al., 2000) as well as having transformed functions through those relations (Smyth et al., 2006). Researchers could take the core procedural property of this task (pairing of stimuli) and develop it elsewhere. For example, participants could be exposed to a fictitious newspaper (or media) advertisement in which an existing Brand A is repeatedly paired with a novel Brand B, and Brand B is then paired with another new Brand C. Alternatively, the respondent-type task could be used to establish a derived transformation of function for three novel brand products (A-B-C) in the laboratory. Thereafter, these (as well as other untrained novel) brands could be offered as promotional “new products” in a student cafeteria, shop or meeting area. Participants’ selection of the brands could then be recorded.

It seems important to note that a significantly larger number of participants passed the derivation test compared to our earlier studies (80% versus 50% and 54%). One promising explanation may reside in the ecological validity of our “marketing” task. In contrast to the highly abstract preparations used in Experiments 1-3 in which arbitrary cartoon characters had to be related to one another using abstract symbols for no apparent reason, Experiment 4 provided participants with a plausible background rationale for learning (i.e., marketing research involving European brand products). This may have helped “contextualize” the study and facilitated the relating of those stimuli. If this assumption is correct then a

potentially interesting question emerges: to what degree does second-order contextual control in the form of “cover stories”, more ecologically valid stimuli or preparations influence the probability of derivation?

On the one hand, our procedures allowed tight experimental control to be exerted over stimulus relating. Participants completed a learning task in which previously unknown stimuli with a single psychological function were related to other abstract stimuli using geometric symbols. In doing so, we sought to minimize the potential for erroneous factors to contaminate the behavioural process under investigation. On the other hand, it seems rather unlikely that this is how people habitually relate stimuli in their everyday lives. Within the wider world stimuli often have multiple functions, elicit responses with meaningful consequences and are related within the boundaries of some wider context. For instance, people may opt for one item over another in a supermarket based on a number of functions (e.g., price, quality, quantity) and the context within which they are related (e.g., whether that person is shopping on a budget or has an abundance of money to spend). Therefore it may be time to explore whether relational performances are facilitated when stimuli and preparations are used that more closely approximate how people typically behave in their everyday lives. Indeed, this form of second-order contextual control could represent a stepping stone towards the eventual goal of demonstrating a transformation of function in applied contexts, stimuli and populations.

Another issue worth noting is that 43% of participants in the fail (derivation) group opted for the correct brand products during the behavioural choice task while 57% of this group selected brands randomly. While 43% correct selection was significantly lower than that observed for the pass group (86% correct selection), it is unlikely to be due to chance factors given the complexity of the relations. Thus, it may be that the 43% of the fail group who selected correctly on the behavioural choice task had started to derive within the testing

context itself and this subsequently influenced their responding on the selection test. For example, they may have responded incorrectly on the first few testing trials, related accurately thereafter and yet still have “failed” the test. If correct, then our mastery criteria on the derivation test may be serving to select out participants who derived *prior* to the testing phase and discard those that derive *within* the test itself. This is broadly consistent with our findings in Experiment 3 (i.e., participants who failed the derivation test before the IAT showed evidence of larger effects relative to those that failed the test after the IAT). If correct, then testing for derivation is not necessary for derived stimulus relations to emerge. Nevertheless, such tests may function as an experimental “context of relating” that increases the probability of derivation taking place – especially for those that have not derived prior to that test (for related findings see Smeets & Barnes-Holmes, 2003). Although speculative, future work could explore this possibility in greater detail.

Finally, while both indirect measures provided evidence for brief and immediate relational responding, the coordination and opposition effects obtained from the IAT were relatively more robust than their affective priming counterparts. In explaining this outcome it may be important to appreciate that, despite its widespread use and apparent utility, the affective priming task is subject to exceptionally low reliability (e.g., Bar-Anan & Nosek, 2012; LeBel & Paunonen, 2011; Uhlmann, Pizarro & Bloom, 2008). Given that the probability of replicating an experimental effect decreases and potential to miss those effects increases with high degrees of measurement error (i.e., low reliability) the above priming effects should be interpreted with caution. Importantly, we are not suggesting that these priming effects are uninformative - they clearly are. In-line with the IAT and IRAP effects obtained in Experiments 1-4, performance on the priming task was driven by a derived transformation of function through the stimulus relations. Rather we argue that future work should replicate our findings using an alternative priming task that is characterized by better

psychometric properties. The Affective Misattribution Procedure (AMP) appears to be one viable candidate (Payne et al., 2005).

Chapter 5: A Derived Transformation of Functions through Comparative Relations as Measured by the IAT and Self-Report Tasks

Across a number of studies, stimulus sets and procedures, psychological functions established for one stimulus have quickly and spontaneously been acquired by other indirectly related stimuli in the absence of pairings, reinforcement or instruction. Indeed in Experiments 1-4 contrasting patterns of evaluative responding have been engineered towards Pokémon characters and fictitious brand products and these responses have been captured using measures such as the IAT, IRAP and Affective Priming. One common thread weaving each of the foregoing experiments together is that only those participants who successfully derive a relation between stimuli (and transformed functions through those relations) respond positively or negative towards previously neutral stimuli.

5.1 Experiment 5

Overall then, a transformation of function in accordance with coordination and opposition relations appears to have important implications for the study of human likes and dislikes. However, RFT argues that entirely different patterns of responding may emerge when other types of derived relations are formed. Imagine, for example, that a person learns to respond appropriately to cues meaning 'More than' and 'Less than'. Thereafter they learn that one stimulus (A) is smaller than a second stimulus (B) which is in turn smaller than a third stimulus (C). This set of directly trained 'Less than' relations (e.g., $A < B$, and $B < C$) may give rise to a number of novel, derived relations between those same stimuli. In other words, participants may respond to B as being more than A and C as more than B (mutual entailment) as well as A being as less than C and C more than A (combinatorial entailment) without any explicit training. If a respondent function is then established for the first stimulus in that relation (A) by directly pairing it with a shock, then the second (B) and third stimuli (C) may come to occasion different responses than the first. Stated more precisely, once a

comparative relation between A, B and C is established, and A becomes a stimulus that predicts mild shock, B may elicit larger fear responses than A while C may elicit even greater levels of fear responding than either A or B (see Dougher et al., 2007 for related findings). Put simply, while psychological functions may be broadly similar for stimuli in a coordination relation, or even reversed for those in an opposition relation, they typically vary in degree rather than kind within a comparative relation.

To date, discriminative (Dymond & Barnes, 1995), consequential (Whelan et al., 2006) and respondent (Dougher et al., 2007) functions have all been transformed through comparative relations in a number of different populations. It remains to be seen whether people will differentially like or dislike stimuli based on their location within a comparative relation. Moreover, the potential utility of self-report and indirect procedures has yet to be fully explored in this domain (although see Bar-Anan & Dahan, in press). With this in mind, Experiment 5 examined whether comparative relations would give rise to evaluative responses that differ not in their direction but rather in their magnitude. Prior to the study, participants were informed that during the experiment they would encounter the names of several “prizes”. They were also informed that they could take one of these prizes home with them at the end of the task. Contextual cue training was then administered in order to establish the relational functions of ‘More than’ for one symbol and ‘Less than’ for another. These cues were used to generate a single comparative relation comprised of five different non-sense words (*Pardal < Zatte < Ettalas < Ciney < Witkap*); the reason why we trained a five- rather than three-member relational network will be explained in the *Method* section. Thereafter, a consequential function was then established for Pardal and Zatte by making access to different quantities of money contingent on their selection. Specifically, a number of trials were presented that allowed participants to increase their overall winnings by picking one of two prizes; Pardal (worth 1 cent) or Zatte (worth 25 cents). Following training, a

derivation test, direct and indirect measures of evaluation and a behavioural choice task were completed.

If a comparative relation is formed as predicted then two outcomes should be evident. First, self-reported ratings should vary according to a stimulus' location within the derived relation, with Ciney evaluated far more positively than either Ettalas or Zatte, and Ettalas evaluated as less positive than Ciney but more so than Zatte. That is, participants should find stimuli that were never paired with money more appetitive than stimuli that reliably and consistently increased their overall prize winnings. Second, this experimentally established history of learning should also be evident when indirect procedures are used to target brief and immediate relational responding. To test this latter assumption, participants were randomly assigned to one of three IATs assessing either Zatte relative to Ettalas; Ettalas relative to Ciney or Ciney relative to Zatte. If the outcomes obtained on the IAT reflect a history of arbitrarily applicable relational responding as suggested, then participants should demonstrate a response bias favouring Ettalas over Zatte, Ciney over Ettalas and Ciney over Zatte in-line with the derived relations. We would also expect participants to consistently select Ciney when offered the opportunity to pick one of three prizes (Zatte, Ettalas or Ciney) in a behavioural choice task administered at the end of the study. In-line with our previous findings, we only expected this pattern of responding to emerge when participants derived the relation between the trained stimuli. Those that fail to do so should not only rate Zatte more positively than Ettalas or Ciney but also favour Zatte over Ettalas or Ciney on the IAT and behavioural choice task.

Finally, the derivation test used in Experiments 2-4 was modified in order to test for mutual (Ettalas > Zatte; Ciney > Ettalas) and combinatorially entailed relations (Ciney > Zatte; Zatte < Ciney) that had not been explicitly trained during the study.

5.2 Method

Participants and Design

Eighty one undergraduates (51 female) ranging from 18 to 34 years ($M = 20$, $SD = 2.8$) participated in exchange for a small sum of money. A $3(IAT) \times 2(Task\ order)$ design was employed with both variables manipulated between-participants. Data from fifteen individuals who failed to achieve the mastery criteria during training were excluded from subsequent analyses.

Materials

Stimuli. Five of the non-sense words used in the previous study served as CSs (*Pardal, Zatte, Ettalas, Ciney* and *Witkap*) while different quantities of money served as USs (1 cent vs. 25 cents). The same arbitrary symbols (i.e., Q and A) were used as contextual cues. All of the CSs were pre-tested and selected on the basis of their neutral ratings and low variability of evaluations.

Indirect procedures. Across the three IATs two of the prize names (*Zatte, Ettalas, or Ciney*) served as one set of category labels and the words “Good” and “Bad” as another. Six positive and six negative adjectives served as one set of attribute stimuli (*wonderful, best, superb, excellent, amazing, great, pleasant, nice* versus *terrible, awful, worse, horrible, nasty, unpleasant, bad, rubbish*) while two prize names served as a second set.

Procedure

Participants were informed that during the study they would encounter the names of several prizes, one of which they could take home with them at the end of the experiment. Thereafter, they were exposed to the following five phases; contextual cue training, relational training, a derivation test, (in)direct measures of evaluation and a behavioural choice task.

Contextual Cue Training

The procedure used to establish the relational functions of ‘More than’ and ‘Less

than' for two arbitrary symbols was similar to that employed previously. On-screen instructions informed participants that the computer would present a series of trials containing two pictures at the top of the screen and two symbols at the bottom of the screen. Their goal was to determine the meaning of the two symbols at the bottom of the screen by using the corrective feedback provided by the computer. Participants were asked to take their time throughout the task and try to respond as accurately as possible. Thereafter the researcher left the room and training began.

Contextual cue training consisted of two different types of trials – those designed to establish the relational function of 'More than' for one symbol and 'Less than' for a second. In either case, two pictures were presented at the top of the screen and two symbols on the bottom (see Figure 5.1). To ensure that quantity - and not some other property - was the dimensional along which the stimuli were related, a red circle was used as a discriminative stimulus for cue selection. Specifically, when presented with (a) an image of a red circle containing many items and (b) an image with a smaller number of items not in a circle, selecting the 'More than' cue was reinforced (i.e., the written feedback "Correct" appeared on-screen for 1000ms). Thereafter, all stimuli would disappear, and following a brief inter-trial interval, the next trial would begin. If participants selected the incorrect ('Less than') symbol when presented with the above images corrective feedback appeared on-screen. In order to remove this feedback and continue with the task the correct response had to be emitted. An entirely opposite pattern of responding was required on trials designed to establish the relational functions of 'Less than' for the second symbol. Specifically, selecting the 'Less than' cue was reinforced when the picture containing the smaller number of items was enclosed in a red circle and the picture containing the larger number of items was not surrounded by a circle. Although contextual cue training started by using stimuli that were formally related to one another (e.g., more ants versus less ants; more apples versus less

apples) the task quickly abstracted this non-arbitrary relation to a purely arbitrary one by including stimuli that were not formally related to one another (e.g., more pentagons versus less dogs; more squares versus less ants). Note that stimulus presentations were counterbalanced such that the same picture sometimes occasioned the selection of the ‘More than’ cue and at other times occasioned the selection of the ‘Less than’ cue. For instance, when presented with a picture of two ants and a red circle surrounding a large number of apples, selecting the ‘More than’ cue was reinforced. However, when presented with two ants and a red circle containing a single bicycle, selecting the ‘Less than’ cue was reinforced.

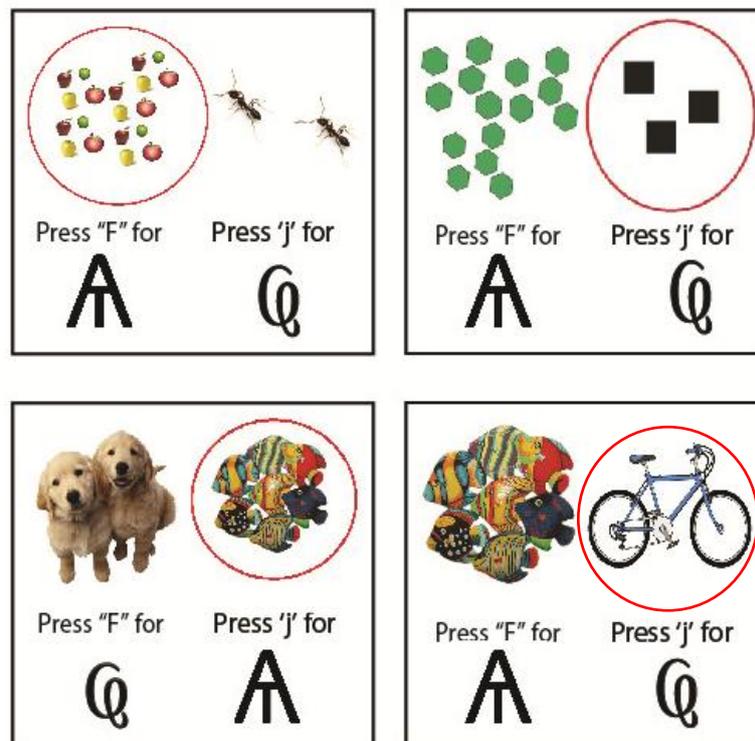


Figure 5.1. Four examples of the ‘More than’ and ‘Less than’ training trials. Each trial consisted of two pictures at the top of the screen and two contextual cues at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in “Correct” being presented in the middle of the screen while selecting the cue deemed incorrect caused “Incorrect” to appear (no feedback was presented during test trials).

Participants were exposed to a minimum of one and a maximum of three blocks of 50 training trials. Within each block the allocation of the two symbols to the lower left and right sides of the screen, as well as presentation of the ‘More than’ and ‘Less than’ trials was varied in a quasi-random order. Progression from training to testing required that participants

respond with 100% accuracy across 20 successive trials, while progression from the test block to relational training required that they respond correctly on at least 20 out of 24 test trials. Failure to do so resulted in re-exposure to another set of training and testing blocks until these mastery criteria were met. Failure to attain criteria following a total of three training and testing blocks resulted in participants being thanked, debriefed and dismissed.

Relational Training

A single comparative relation consisting of five fictitious “prizes” was generated using a similar protocol as before (*Pardal < Zatte < Ettalas < Ciney < Witkap*) (see Figure 5.2). Training consisted of four phases, each with a minimum of one and a maximum of three blocks of 50 trials. Participants were informed that during this section of the task they would be presented with the names of several prizes they could potentially win at the end of the experiment. On each trial, the names of two prizes would be presented on the upper left and right sides of the screen while the two symbols they had previously encountered would appear at the bottom of the screen. Their task was to determine the relationship between the two prizes using the corrective feedback provided by the computer.

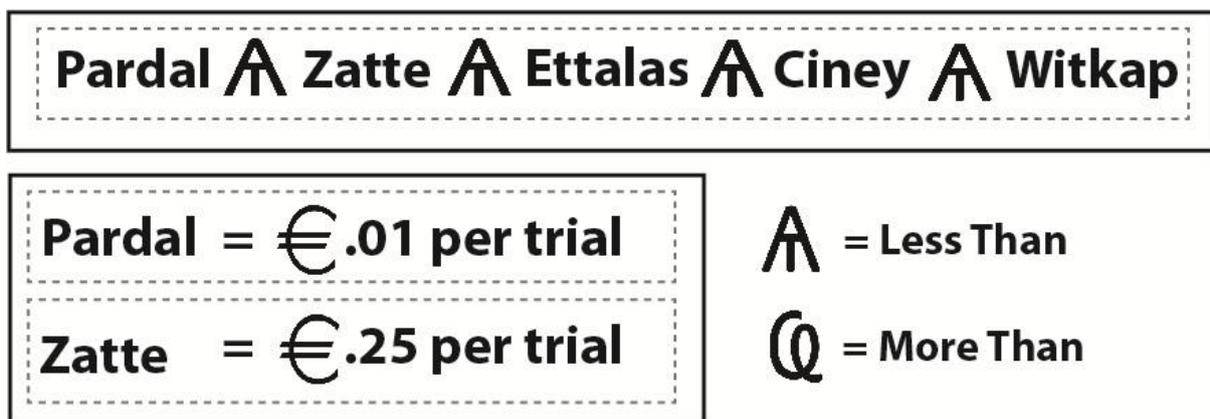


Figure 5.2. Schematic representation of the comparative relation established in Experiment 5. Five fictitious brand products participated in the relation (*Pardal-LessThan-Zatte-LessThan-Ettalas-LessThan-Ciney-LessThan-Witkap*). A consequential function was then established for Pardal by pairing it with repeated access to €0.01 and Zatte with €0.25 per trial.

During the first phase of training, three stimulus relations were established by differentially reinforcing the selection of one of the two contextual cues in the presence of a

specific stimulus combination (*Pardal* < *Zatte*; *Zatte* > *Pardal*; *Zatte* < *Ettalas*). For example, when a red circle containing *Pardal* was presented with *Zatte* or a red circle containing *Zatte* was presented with *Ettalas*, selecting the ‘Less than’ cue resulted in “Correct” appearing on-screen for 1000ms, followed by a brief inter-trial interval and the following trial. Emitting an incorrect response - such as selecting ‘More than’ in the presence of the above stimuli - caused error feedback to appear. Progression to the next trial was made contingent on selecting the correct (‘Less than’) cue. In contrast, when a red circle containing *Zatte* was presented together with *Pardal*, selecting the ‘More than’ cue was reinforced and the ‘Less than’ cue punished (see Figure 5.3). Finally, in addition to the above relations, and to ensure that the ‘More than’ and ‘Less than’ functions remained salient throughout the task, a number of contextual training trials were also interspersed within each block.

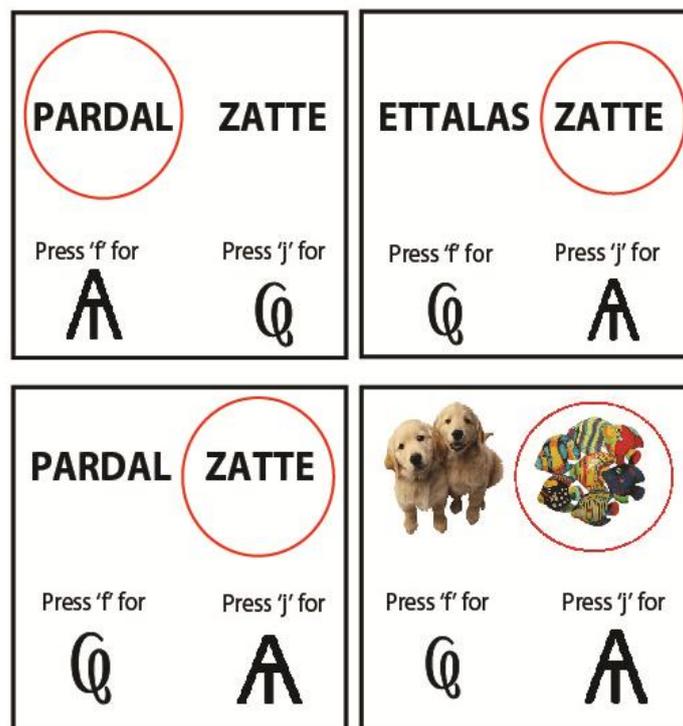


Figure 5.3. Examples of the four trials involved in the first phase of (comparative) relational training. Each trial displayed two prize names at the top of the screen and the two contextual cues at the bottom of the screen.

The second phase of training was identical to the first with the exception that three different relations were established (*Ettalas* < *Ciney*; *Ciney* < *Witkap*; *Witkap* > *Ciney*) while

the third phase exposed participants to all six relations generated during the task. An evaluative function was then established for Pardal and Zatte by making access to different amounts of money contingent on their selection. On-screen instructions informed participants that in the next part of the study they could win a small sum of money. Specifically, two of the prizes they had just encountered would be presented on the lower left and right corners of the screen and they could add one of these prizes to their overall winnings by clicking on it with the mouse. Fifteen trials were then presented, each with a label at the top of the screen displaying the “number of opportunities remaining”, a second label underneath stating their overall prize winnings and the two prizes “Pardal” and “Zatte” at the bottom of the screen. Selecting Zatte on any given trial increased participants’ total winnings by 25 cents while choosing Pardal only added one cent to that amount. Given that participants were provided with an initial sum of 75 cents on the first training trial they could win a maximum of €4.50 by consistently choosing Zatte (or a minimum of one euro for consistently selecting Pardal) across each of the fifteen trials.

Test for Derived Relational Responding

To determine whether a comparative relation was formed as predicted (and the evaluative function established for Zatte was transformed through that relation to Ettalas and Ciney), participants were tested for mutual and combinatorial entailment. On each trial, two prizes were presented along with the ‘More than’ and ‘Less than’ cues and participants were asked to “*click on the symbol that describes the relationship between the two prizes*”.

Selecting a contextual cue removed all stimuli from the screen, onset a 500ms inter-trial interval and the next trial. Testing consisted of twelve trials, four of which presented Zatte and Ettalas together, another four presenting Ettalas and Ciney while the final four presented Zatte and Ciney together. Critically, at no time was feedback provided for any response emitted during this task.

Responding in accordance with the predicted comparative relation required participants to select the ‘More than’ cue when Ciney was surrounded by a circle and presented with either Ettalas or Zatte; choosing the ‘Less than’ cue when Zatte or Ettalas was surrounded by a circle and presented with Ciney or the ‘More than’ cue when Ettalas was surrounded by a circle and presented with Zatte. To ensure that the critical elements in the network were only trained in one direction, the “endpoint” pairs Pardal < Zatte or Witkap > Ciney were not tested since the selection of either Pardal or Witkap was always reinforced in both directions (e.g., participants were directly trained that Pardal < Zatte and Zatte > Pardal as well as Ciney < Witkap and Witkap > Ciney). A minimum of 10 out of 12 correct responses was required to pass the test and those who did not meet this criterion were defined as having failed the test.

Indirect Procedure

Participants were randomly assigned to complete one of three different IATs in order to demonstrate that comparative relations give rise to a relativistic set of brief and immediate relational responses. On the first IAT evaluative responding towards Zatte was assessed relative to Ettalas; the second IAT examined Ciney relative to Ettalas while the third IAT targeted Ciney relative to Zatte. Across all three variants of the task two of the previously encountered prizes (Zatte, Ettalas or Ciney) functioned as one set of category labels and attribute stimuli. The words “Good” and “Bad” as well as six positive and negative adjectives served as a second set of category and attribute stimuli. If comparative relations were established in-line with prior training then response latencies should be shorter during consistent relative to inconsistent IAT trials. For instance, on an IAT assessing Zatte relative to Ettalas, participants should be quicker in pairing Ettalas with positive stimuli than Zatte. However, when Ettalas is assessed relative to Ciney, participants should be quicker in pairing

Ciney with positive stimuli than Ettalas (a similar response bias in relating Ciney with positive stimuli should also be evident on a Ciney-Zatte IAT).

Direct Procedures

A similar set of stimulus rating, contextual cue meaning and demand compliance tasks were employed as in previous studies.

Behavioural choice task. Following the various measures of evaluation, participants were presented with three small boxes that were identical in size, shape and colour. Each box was labelled with one of three prize names (Zatte, Ettalas or Ciney). Participants were offered the opportunity to select one of the prizes to add to their overall winnings as a final “thank you” for taking part in the experiment. After participants made their selection they were thanked, debriefed and dismissed.

5.3 Results

Data Preparation

Contextual cue meaning. Of the current sample sixty three participants (95%) reported the relational functions of the two contextual cues in-line with experimental expectations. On the one hand, thirty three participants (53%) rated the ‘More than’ cue as meaning “more than”; sixteen participants (25%) rated it as meaning “greater than” while another fourteen (22%) used terms such as “larger than”, or “many”. On the other hand, forty five participants (71%) rated the ‘Less than’ cue as “less than”; while the remaining eighteen (29%) used terms such as “smaller than”, “fewer than” or “lower than”. Data for the two participants who reported incorrect relational functions were removed prior to analysis. Reanalyzing the data with these participants included did not change any of the statistical conclusions reported below.

Demand compliance. Only one participant reported that they intentionally responded to the prizes in-line with the presumed expectations of the experimenter. Excluding these data did not impact any of the obtained effects outlined below.

Preliminary analyses. Counterbalancing the order of direct and indirect tasks as well as IAT test blocks produced no significant effects. As such, analyses were collapsed across these two method factors. Fourteen participants (22%) failed the derivation test while forty nine (78%) passed the test.

Self-Reported Ratings

Mean likeability scores for Zatte, Ettalas, and Ciney as a function of derivation test performance are presented in Figure 5.4. As can be seen in the graph, a relativistic and graded pattern of self-reported ratings emerged based on the location of a stimulus within the comparative relation. Despite being the only prize that predicted access to money, participants rated Zatte as less positive than Ettalas while Ettalas in turn was rated as less positive than Ciney. Mirroring our previous findings, evaluative responding only emerged when participants successfully passed the derivation test. Failure to do so resulted in largely ambivalent responding towards each of the three prizes.

When submitted to a 3(*Prize*) x 2(*Derivation test*) mixed-model ANOVA, a main effect was obtained for prize, $F(2, 61) = 4.3, p = .02, \eta^2_{\text{Partial}} = .07$, as well as a two-way interaction between prize and test performance, $F(2, 61) = 8.5, p = .001, \eta^2_{\text{Partial}} = .12$. To specify this interaction, self-reported ratings for those who passed the derivation test were analyzed separately from those who failed to do so. With respect to the pass group, ratings of the three prizes differed significantly from one another, $F(2, 48) = 45.9, p = .001, \eta^2_{\text{Partial}} = .49$, with simple comparisons indicating that Ettalas ($M = 1.0, SE = .27$) was rated more positively than Zatte ($M = .27, SE = .29$), ($p = .004$) while Ciney ($M = 2.9, SE = .2$) was rated more positively than either Zatte ($p = .001$) or Ettalas ($p = .001$). While positivity scores for

Ettalas ($p = .001$) and Ciney ($p = .001$) differed significantly from zero, participants rated Zatte (i.e., the stimulus directly paired with money) as neutral ($p = .4$). In direct contrast, participants who failed the derivation test did not rate any of the three stimuli as being positive or negative, $F(2, 13) = .1, p = .9$, with Zatte ($M = 1.0, SE = .67$), Ettalas ($M = .81, SE = .73$) and Ciney ($M = .53, SE = .86$) each eliciting largely ambivalent responses (all $ps > .2$).

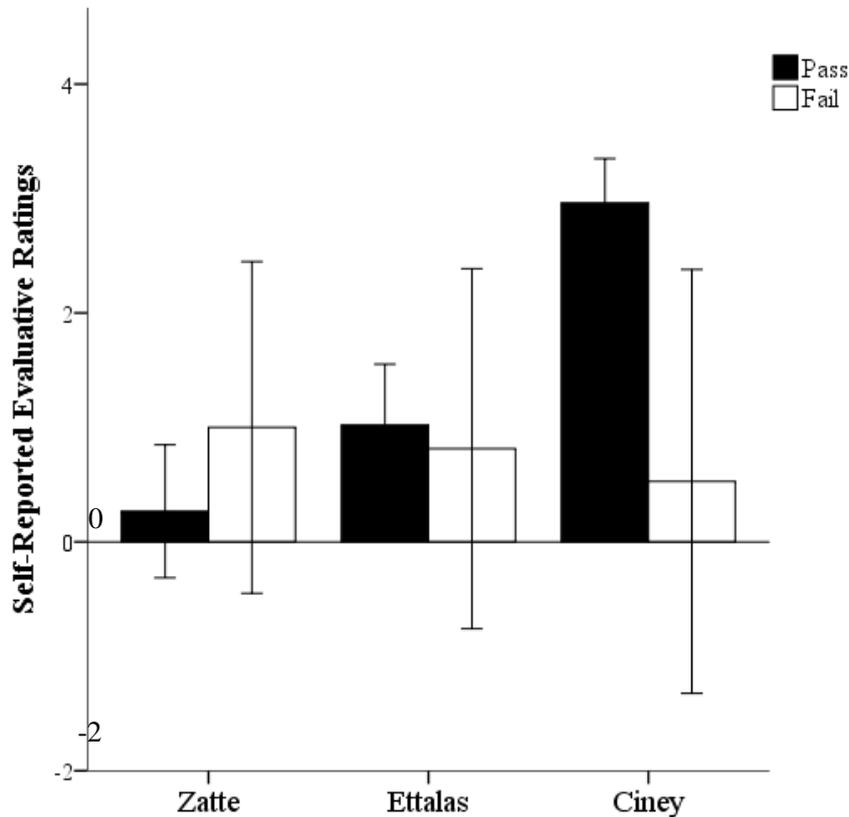


Figure 5.4. Mean likeability scores for Zatte, Ettalas and Ciney as a function of derivation test performance (pass vs. fail). Error bars indicate standard errors.

Indirect Procedure

Data from the three IATs were transformed using the D6 algorithm and scored so that positive values reflected a response bias favouring the prize further along the comparative relation. For instance, when an IAT was used to examine responding towards Zatte relative to Ettalas, positive scores reflected a bias favouring Ettalas over Zatte. Nonetheless, when the task assessed responding towards Ettalas relative to Ciney, a positive score indicated a bias for Ciney over Ettalas (a negative value for any of the three IATs indicated a reversed pattern

of responding). Given the restricted size of the fail group meaningful analysis across the various experimental conditions was not possible. Consequently, only IAT data from the pass group is analysed below.

To determine whether brief and immediate relational responding differed as a function of the stimulus relation assessed, scores from three IATs were submitted to a one-way between groups ANOVA (no effect was obtained for IAT type; $F(2, 48) = .26, p = .78$). Consistent with our predictions, follow-up one sample t-tests revealed that participants who passed the derivation test demonstrated a response bias favouring Ettalas relative to Zatte ($M = .24 SE = .1$), $t(14) = 2.3, p = .04$, Ciney relative to Ettalas, ($M = .29 SE = .07$), $t(11) = 4.1, p = .002$, and Ciney relative to Zatte ($M = .19 SE = .09$), $t(21) = 2.2, p = .04$ (see Figure 5.5).

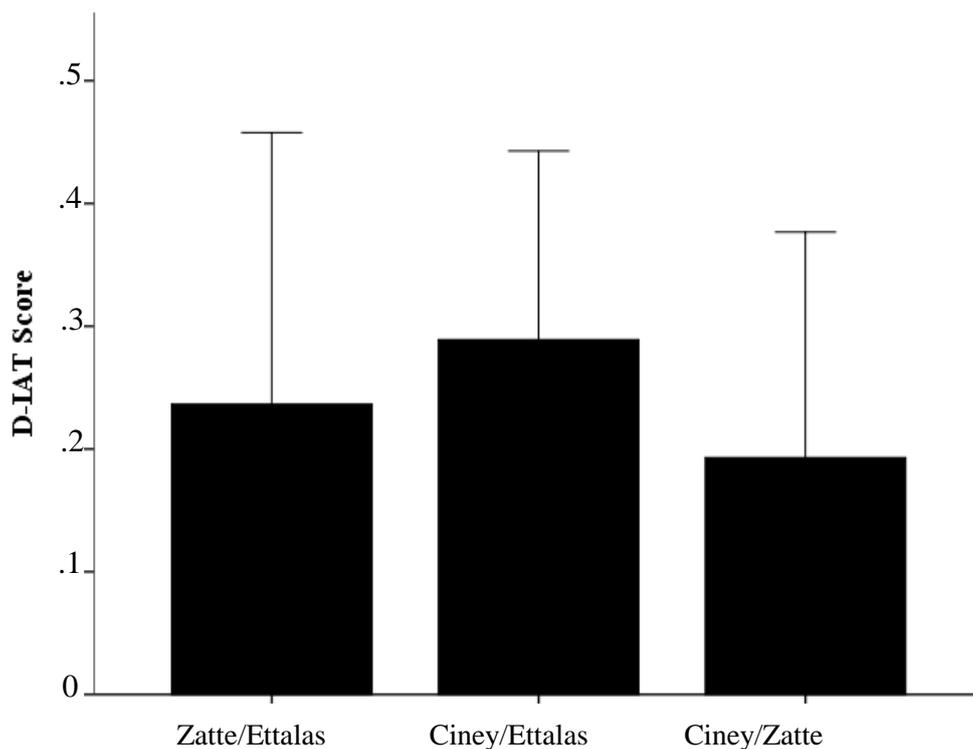


Figure 5.5. Mean D-IAT scores for Zatte relative to Ettalas, Ciney relative to Ettalas and Ciney relative to Zatte. A positive score indicates a response bias for the stimulus further along the comparative relation. Error bars represent standard errors.

Behavioural Choice Task

A correct response was defined as the selection of Ciney while an incorrect response as the selection of either Zatte or Ettalas when provided with the option to choose from the

three prizes. Eighty eight percent of participants who passed the derivation test selected the correct option while none of the fail group made a comparable choice.

Discussion

The current experiment indicates that comparative relations give rise to a unique pattern of evaluative responding that is inherently relativistic in nature. Using broadly similar procedures as before, a five member stimulus relation was generated (*Pardal* < *Zatte* < *Ettalas* < *Ciney* < *Witkap*) and consequential functions established for the first (*Pardal*) and second (*Zatte*) stimuli by making access to different amounts of money contingent on their selection (1 cent per trial versus 25 cents per trial). When direct and indirect measures were used to assess *Zatte*, *Ettalas* and *Ciney*, the stimuli were evaluated differently depending on their location within the derived relation. Despite being the only prize that consistently and reliably increased the amount of money that participants could win, *Zatte* was liked less than *Ettalas* while *Ettalas* was liked less than *Ciney*. This comparative relation was also evident on both the IAT and behavioural choice tasks, with participants favouring *Ettalas* relative to *Zatte*, *Ciney* relative to *Ettalas* and *Ciney* relative to *Zatte* on the former and *Ciney* on the latter. Consistent with Experiments 2-4, evaluative effects were only observed when participants showed evidence of having formed derived stimulus relations. Those that failed to do so produced no evaluative responses towards any of the prizes and selected randomly from the available options on a behavioural choice task.

5.4 Experiment 6

The above data extend the remit of our relational account from coordination and opposition relations to those that involve comparison. Nevertheless, several methodological issues still need to be addressed. In Experiment 5, for example, assignment to the three IATs was manipulated across rather than within participants. This strategy introduces the possibility that the relational outcomes obtained on the IAT do not reflect genuine

comparative relating at the level of the individual but only emerge when different groups are assessed. In order to control for this criticism, we replicated the previous experiment while ensuring that participants completed each of the three IATs in a random order. At the same time, Experiment 3 demonstrated that a transformation of function can occur regardless of whether a derivation test is administered before or after the critical measures of evaluation. It remains to be seen whether this also holds true for comparative relations. Therefore in Experiment 6 the presentation order of the derivation test was manipulated such that half of the participants completed the test immediately after relational training while the other half encountered it at the end of the study. Likewise, the order of the behavioural choice task was also counterbalanced to ensure that exposure to the measures of evaluation did not influence the “prize” participants chose to take home with them. Half of the participants were therefore asked to pick a prize immediately after relational training while the other half completed the task at the end of the study.

Finally, while significant IAT effects were observed in Experiment 5 they were relatively less robust than those obtained previously. One possibility is that the evaluative functions established for Pardal and Zatte were relatively “weak” (i.e., participants may not have cared much for a stimulus that increased their winnings by a mere 25 cents). As such, we increased the value of Zatte from 25 cents to 1 euro per trial while maintaining the value of Pardal at 1 cent.

5.5 Method

Participants and Design

Forty three undergraduates (35 female) ranging from 18 to 33 years ($M = 21.9$, $SD = 4.6$) participated in exchange for a €5 payment. A $3(IAT) \times 2(Derivation\ test\ performance) \times 2(Derivation\ test\ time)$ design was employed with the latter variable manipulated between-participants. Three additional method factors were also manipulated: the order in which

participant completed the direct and indirect tasks, IAT block order and whether the behavioural choice task was administered before or after the other measures of evaluation. Data from six individuals who failed to achieve the mastery criteria during training were excluded from analyses.

Procedure

The procedures used to train and test derived comparative relating were similar to those employed in the previous study - with four notable exceptions. First, brief and immediate relational responding was assessed within rather than between participants. To minimize the potential for fatigue, a shortened five-block version of the IAT was used with 20 trials per practice block and 30 per test block. Second, and in order to ensure that exposure to a derivation test was not a pre-requisite for evaluative responding, the presentation order of the test was once again manipulated. Third, exposure to the behavioural choice task was also counterbalanced across participants such that half of the participants were asked to pick a prize immediately after relational training while the other half completed the task at the end of the study. Finally, the assumed appetitive functions of Zatte were increased by changing its monetary value from 25 cents to 1 euro per trial (the value of Pardal remained unchanged at 1 cent). In doing so, we sought to increase the evaluative response elicited by this stimulus as well as those indirectly related to it via derivation.

5.6 Results

Data Preparation

Contextual cue meaning. Of the current sample thirty-seven participants (100%) reported the relational functions of the contextual cues in-line with experimental expectations. On the one hand, twenty two participants (59%) rated the ‘More than’ cue as meaning “more than”; eight participants (22%) rated it as meaning “greater than” while the remaining seven (19%) used terms such as “larger” or “bigger than”. On the other hand,

thirty two participants (86%) rated the ‘Less than’ cue as “less than”; while the remaining five (14%) used terms such as “smaller” or “lower than”.

Demand compliance. Only one participant reported that they intentionally responded to stimuli in-line with their assumed expectations of the experimenter. Excluding the data for this participant did not impact any of the obtained effects outlined below.

Preliminary analyses. Counterbalancing the order of direct and indirect tasks, derivation test time, IAT test blocks and exposure to the behavioural choice task produced no significant effects. As such, analyses were collapsed across these various factors. Thirty two participants (84%) passed the derivation test while five (16%) failed the test. Given the restricted size of the fail group, meaningful analysis across the various experimental conditions was not possible. Consequently, only data from the pass group is analysed below.

Self-Reported Ratings

Mean likeability scores for Zatte, Ettalas, and Ciney are presented in Figure 5.5. Similar to before, a relativistic and graded pattern of evaluative responding was obtained, with participants reporting Zatte as less positive than Ettalas and Ettalas less positive than Ciney. When self-reported ratings were submitted to a one-way repeated measures ANOVA, a main effect was obtained for prize, $F(2, 30) = 58.1, p = .001, \eta^2_{\text{Partial}} = .66$, indicating that ratings of the three stimuli differed significantly from one another. Simple comparisons revealed that Ettalas ($M = .83, SE = .28$) was evaluated as more positive than Zatte ($M = -.24, SE = .33$), ($p = .005$) while Ciney ($M = 3.24, SE = .19$) was evaluated more positively than either Zatte ($p = .001$) or Ettalas ($p = .001$). While ratings for Ettalas ($p = .006$) and Ciney ($p = .001$) differed significantly from zero, participants rated Zatte (i.e., the only stimulus that predicted access to money) as neutral ($p = .5$).

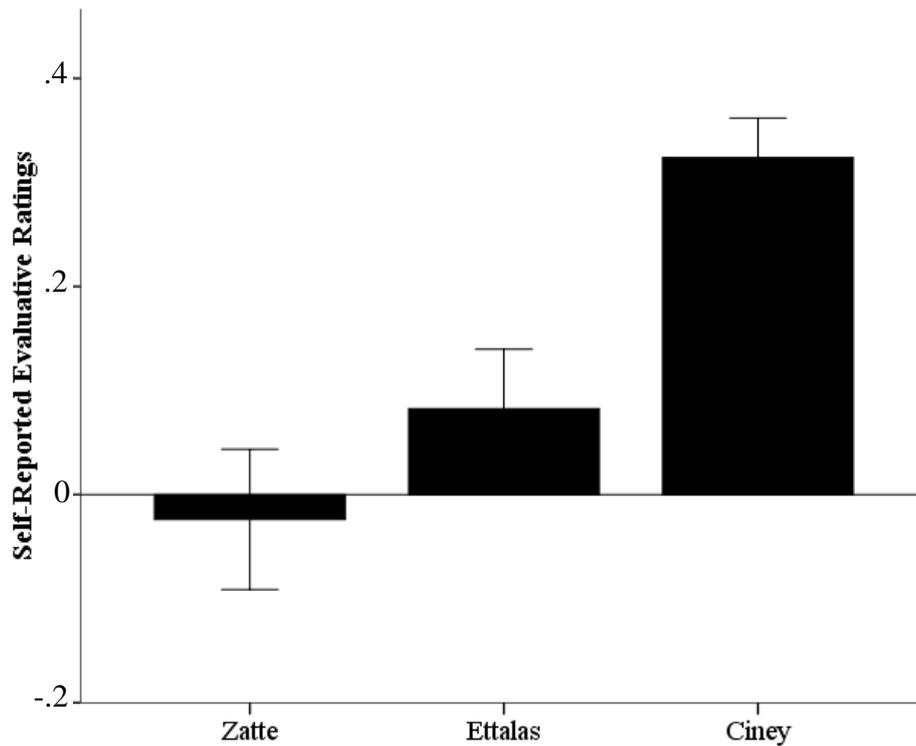


Figure 5.6. Mean likeability scores for Zatte, Ettalas and Ciney. Error bars indicate standard errors.

Indirect Procedure

Data from the three IATs were transformed using the D6 algorithm and scored in a similar manner to Experiment 5. To determine whether brief and immediate relational responding differed as a function of the stimulus relation assessed, IAT scores were submitted to a one-way repeated measures ANOVA (no main effect for IAT type was obtained; $F(2, 30) = .68, p = .5$). A series of one sample t-tests revealed that participants who passed the derivation test preferred Ettalas relative to Zatte ($M = .17, SE = .06, t(30) = 2.8, p = .01$, Ciney relative to Ettalas, ($M = .23, SE = .07, t(30) = 3.5, p = .001$, and Ciney relative to Zatte ($M = .27, SE = .05, t(30) = 4.9, p = .001$ in-line with a derived transformation of function through the comparative relations (see Figure 5.7).

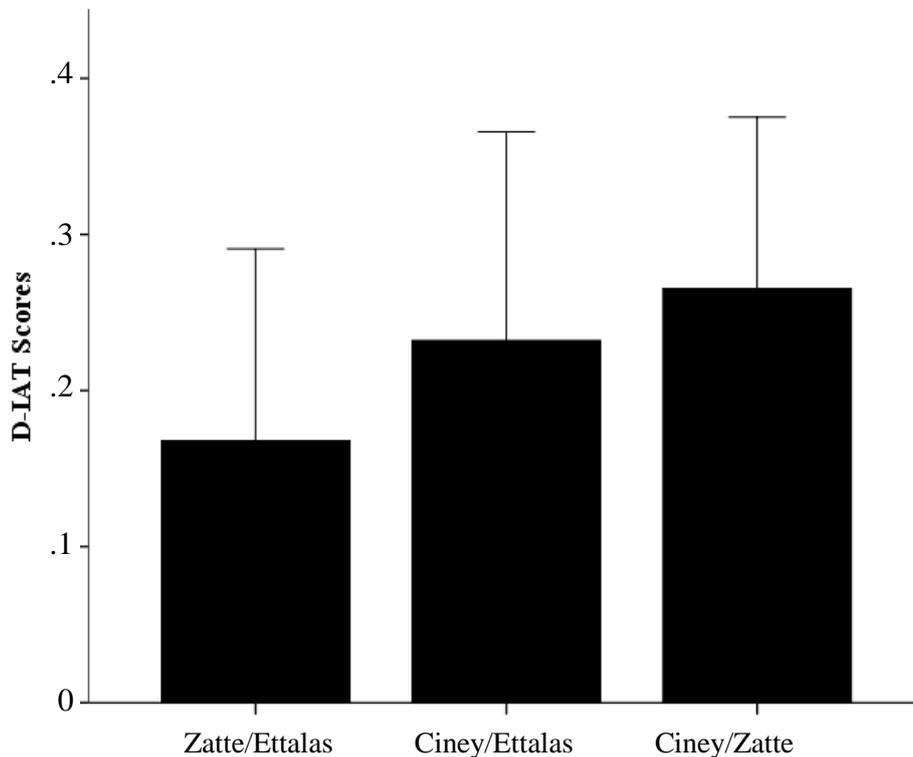


Figure 5.7. Mean D-IAT scores for Zatte relative to Ettalas, Ciney relative to Ciney and Ciney relative to Zatte. A positive score indicates a response bias for the stimulus further along the comparative relation. Error bars represent standard errors.

Behavioural Choice Task

Thirty out of thirty-seven participants (87%) who passed the derivation test selected the correct prize while no participant in the fail group made a comparable choice.

Discussion

The above findings indicate that forming comparative relations among stimuli results in differential evaluative responses to those stimuli as a function of their respective locations within that relation. When participants learned that five fictitious prizes were more or less than one another, and that one of those prizes (Zatte) was worth a euro, they responded in a similar fashion to Experiment 5. In particular, relational performances that were never directly trained but nonetheless relationally coherent with prior learning were observed. Despite being the only stimulus that consistently predicted access to money, Zatte was rated less positively than Ettalas while Ettalas was rated less positively than Ciney. These relational effects were also evident when participants were exposed to three IATs and a behavioural

choice task, with a response bias favouring Ettalas relative to Zatte, Ciney relative to Ettalas and Ciney relative to Zatte on the former (indirect) measure and Ciney on the latter (direct) measure. Importantly, the current work extends beyond our previous findings given that direct and indirect effects were obtained within (rather than between) subjects and regardless of whether a derivation test was encountered before or after the various measures of evaluation. Thus it appears that contextual cues such as ‘More than’ or ‘Less than’ can be used to weave derived relations between and among stimuli, and by implication, transform the functions of those stimuli accordingly.

5.7 General Discussion

Relational Frame Theory argues that (evaluative) functions may be transformed through a wide variety of stimulus relations and that this behavioural process is under the control of contextual cues present in the environment. Although Experiments 1-4 have modelled a transformation of functions through coordination and opposition relations, RFT predicts that an entirely different pattern of responding should be obtained when a set of ‘More than’ and ‘Less than’ relations are engineered. Specifically, and given that comparative relations involve responding to one event in terms another along some quantitative or qualitative dimension, a graded or relativistic pattern of responding should be evident, with a stimulus evaluated according to its location within a relation. In other words, comparative relations should modify the functions of related stimuli such that they occasion evaluative responses that differ in degrees rather than kind.

In order to test this assumption, the relational functions of ‘More than’ and ‘Less than’ were established for two arbitrary symbols, and these contextual cues were used to generate a relational network comprised of five fictitious prizes (*Pardal* < *Zatte* < *Ettalas* < *Ciney* < *Witkap*). Thereafter a consequential function was established by pairing a number of those stimuli (*Pardal* and *Zatte*) with access to different quantities of money. Across two

separate studies we found that the evaluative functions acquired by the prizes were governed by the specified and derived relations established during training. On the stimulus rating task, for instance, participants responded to Zatte as less positive than either Ettalas or Ciney, while Ciney was rated more positively than either Ettalas or Zatte. These relational effects were also evident when three different IATs were administered, with participants favouring Ettalas relative to Zatte, Ciney relative to Ettalas and Ciney relative to Zatte. Furthermore, when given a choice to take Zatte, Ettalas or Ciney home with them at the end of the study, a majority of participants opted for the prize labelled Ciney.

When taken together, the above findings provide the first empirical demonstration of a transformation of evaluative functions in accordance with ‘More than’ and ‘Less than’ relations (as measured by direct and indirect procedures). Indeed, to our knowledge, only a single EC study has explored changes in evaluative responding based on such relations (Bar-Anan & Dahan, in press) and the authors focused solely on contiguous pairings of stimuli measured via self-report procedures. The current work expands on these early findings and offers an explanation for why people like a stimulus that was never paired with appetitive events *even more* than a stimulus that was paired with such events in the past. Put another way, our findings indicate that derived stimulus relations - whether based on coordination, opposition or comparison - play a key role in shaping our evaluative responses to stimuli in the environment.

Perhaps the most important point to appreciate in Experiments 5-6 is that Zatte was the only prize encountered during testing that previously predicted access to money and yet it was evaluated as the least attractive of the three available stimuli. Moreover, Ciney was never paired with money or even Zatte at any point during training yet it was evaluated as more positive than either of the other two prizes. Given the unidirectional training procedure adopted throughout (see Smyth et al., 2006), this behaviour cannot be traced back to a history

of direct contingency control but appears to represent an instance of arbitrarily applicable relational responding. This argument gains support from the fact that the two symbols and arbitrary stimuli used in the current thesis elicited radically different outcomes as a function of the contextual cue and relational training received. In Experiments 1-4, for example, participants responded to \mathbb{Q} and \mathbb{A} as if they meant 'Same' and 'Opposite' while in Experiments 5-6 they behaved as if those same symbols meant 'More than' and 'Less than'. Likewise, Zatte, Ettalas and Ciney occasioned evaluative responses that varied in their absolute direction during Experiment 4 and merely in their magnitude in Experiments 5-6. In other words, there was nothing intrinsically appetitive or aversive about the above non-sense words or symbols prior to training; rather they appear to have acquired their functions directly or indirectly via a generalised type of operant learning that is under both antecedent and consequential stimulus control.

Nevertheless, it is worth noting that Experiments 5-6 primarily focused on a single type of contextual cue ('Less than') to establish the relation and employed a presumably appetitive stimulus (money) to generate the functions of stimuli in that network. In providing a more sophisticated analysis of evaluative responding through comparative relations, future work could expose one half of participants to $A > B > C$ and the other half to $A < B < C$ training. Thereafter, evaluative functions could be established for the A stimulus by pairing it with either an appetitive (money) or aversive (shock) stimulus. From an RFT perspective, this experimental design would enable the magnitude and direction of evaluative responding to be independently manipulated. For instance, we would anticipate that participants exposed to $A > B > C$ training would evaluate the final stimulus in the relation (C) as being less positive (or negative) than B and B less so than A. At the same time, participants exposed to $A < B < C$ should show precisely the opposite pattern of responding - with C evaluated as more positive (or negative) than B and B more so than A. A second but equally interesting

possibility is that the A_{shock} stimulus could be evaluated more positively than the A_{money} stimulus given that a small shock is better than a large shock while a small sum of money is worse than a large sum of money. The current data seem to provide tentative support for this claim given that Zatte - despite repeatedly being paired with money - was evaluated as neutral rather than positive. Put another way, an otherwise appetitive stimulus may elicit evaluative responses that are neutral (or even negative) when it is comparatively framed as the least appetitive stimulus in a relational network.

On balance, while contextual cue training served to establish one symbol as meaning “More than” and a second as “Less than” it failed to do so in a sophisticated fashion. Although the cues occasioned responding to one stimulus in terms of its difference from another along some quantitative dimension, they did not specify the precise nature of this difference. For instance, participants did not know if Ettalas was twice, five times or a hundred times the value of Zatte nor the extent to which Ciney differed from Ettalas. In a recent study, Vitale and colleagues (2008) found that participants were more accurate when they had to respond to specified ($A < B$; $B < C$) versus unspecified ($C > B$; $B < A$) comparative relations. They argued that adults arrive at the laboratory with a protracted history of deriving specified relations and a relatively limited history of deriving unspecified relations. Moreover, when “unspecified relational difficulties are encountered in the natural language environment, adults are likely to seek clarification through additional information” (p.385). While the relation between Zatte, Ettalas and Ciney was specified in Experiments 5 and 6 (i.e., $Zatte < Ettalas < Ciney$) it may be case that further specification (e.g., *Zatte is ten times smaller than Ettalas and Ettalas is ten times smaller than Ciney*) could occasion more robust direct and indirect evaluative effects. With this in mind, future work could modify our cue training protocol so that comparative relational functions are specified to a greater degree. This may require that the pictures used to establish the C_{rels} vary consistently along a

given dimension (e.g., ten ants versus one apple, ten apples versus one pentagon and ten dogs versus one square). Likewise, a single IRAP targeting the relation between Zatte/Ettalas, Ettalas/Ciney and Ciney/Zatte may provide a more effective means to capture brief and immediate relational responding than a series of IATs manipulated between or within participants (see Power, Barnes-Holmes, Barnes-Holmes & Stewart, 2009).

Finally, while we have restricted our analysis of comparative relating to the realm of evaluative responding, this family of relational frames may play a defining role in other psychological domains. For instance, comparative relations may underpin “contrast effects” in social judgements (Schwarz, Münkkel & Hippler, 1990) and evaluations of physical attractiveness (Kenrick & Gutierrez, 1980) not to mention attitudes (Bar-Anan & Dahan, in press) and stereotyping (Manis, Nelson, & Shedler, 1988) (see also Summerville & Roese, 2008). Likewise, approaching the experimental analysis of transitive inference (Munelly & Dymond, 2010) and the indirect acquisition of fear responding (Dougher et al., 2007) from this relational perspective may also provide novel insight into these (and related) phenomena.

Chapter 6: General Discussion

6.1 Overview

The current thesis set out to investigate whether the behavioural process of derived stimulus relating could facilitate a better understanding of human likes and dislikes than direct contingency accounts alone. Consistent with this notion, our work suggests that a sophisticated experimental analysis of evaluative responding cannot focus solely on direct stimulus pairings – at least where verbally trained humans are concerned. Across a series of experiments and in the absence of co-occurrence, reinforcement or instruction, stimuli spontaneously acquired psychological functions by participating in derived coordination, opposition and comparative relations. Indeed, by exerting fine-grained contextual control over how Pokémon characters, fictitious brand products or potential prizes were related to one another, we systematically manipulated the direction and magnitude of ensuing responses on both direct and indirect procedures. In the following chapter, we review this research in detail and consider its conceptual and theoretical implications for both mechanistic and functional researchers. Thereafter, we outline how our findings may inform future research on both evaluative responding and implicit cognition more generally.

6.2 A Functional Approach to Evaluative Responding

Throughout much of the past century learning psychologists have focused their theoretical, methodological and empirical attention on a seemingly simple question: how do people come to like or dislike stimuli in the environment? More often than not, the answer to this conundrum has been pursued by researchers subscribing to a mechanistic world-view wherein mental processes and conditions are argued to mediate between environment and behaviour (Chapter 1). From Monet and Mozart to our favourite foods, films and friends, humans are argued to acquire a vast repository of preferences, attitudes or evaluations that shape their behaviour across a variety of domains. When approached from this perspective,

the empirical agenda quickly shifts to identifying the determinants of these mental constructs and their mechanisms of change. Evaluative conditioning has attracted considerable attention in this regard. Over the last four decades this direct contingency approach has sparked a burgeoning industry of mental theorizing and methodological developments united under one common assumption: the direct pairing of stimuli in space and time represents an important avenue for generating and modifying evaluative responses.

Parallel to these developments, a novel intellectual tradition known as contextual behavioural science (CBS) has taken root and sought to understand, predict and influence human behaviours such as evaluative responding. Equipped with alternative philosophical (functional contextualism) and theoretical frameworks (Relational Frame Theory), contextual behavioural scientists have substituted the mechanistic approach for a purely functional one based on the evolutionary notion of natural selection. At the core of this tradition is an extensive literature indicating that humans behave in a seemingly unique manner that is largely absent elsewhere in the animal kingdom. Whereas many species are shackled to direct contingency learning based on (a) the physical or formal properties of the related stimuli and (b) specific temporal parameters of those relations, humans appear to stand alone in their ability to derive novel, untrained relations between and among different stimuli (i.e., demonstrate evidence for arbitrarily applicable relational responding). With respect to the current thesis, RFT suggests that this ability to respond in an arbitrarily applicable fashion is learned early on in our development and forever changes how we interact with the world (and by implication how we come to like and dislike stimuli). Put another way, the ability to derive relations between one stimulus and another appears to impact upon every other known behavioural principle. For instance, contiguous presentations of stimuli may no longer simply involve respondent learning - rather these stimuli may also be knitted together into derived relations without any training or instruction to do so (Leader et al., 1996; Smeets et al., 1997;

Smyth et al., 2006). Likewise, and following a history of relational learning, an associatively conditioned appetitive CS can immediately be transformed into an aversive CS by participating in an appropriate relational frame (e.g., Whelan & Barnes-Holmes, 2004).

According to RFT, direct contingency accounts that restrict their analysis to stimulus pairings constitute an important first step towards understanding human likes and dislikes – but a first step nonetheless. Although these models have offered many valuable insights they do not explain why people may approach, avoid and respond with fear, disgust or pleasure when faced with entirely novel stimuli that have never predicted or even been related with psychological events in the past. At the same time, they also seemingly fail to identify when, how and why these patterns of derived stimulus relating will be emitted across a spectrum of direct and indirect procedures. The present work constitutes the first systematic attempt to address both of these questions as they apply to human evaluative responding.

6.3 Summary of the current research. Experiment 1 (Chapter 2) demonstrated that a stimulus could be liked or disliked – not based on its co-occurrence with another stimulus - but rather the relation established between those stimuli by a contextual cue. Four mutually entailed relations were formed between Pokémon characters and emotional images using two laboratory-induced cues meaning either ‘Same’ or ‘Opposite’ (i.e., *Pokémon1-Same-Positive*; *Pokémon2-Opposite-Positive*; *Pokémon3-Same-Negative*; *Pokémon4-Opposite-Negative*). Evaluative responding towards the various Pokémon was then assessed using a number of direct and indirect (IAT) procedures. Consistent with our predictions, ratings varied according to what cue governed the stimulus relation during training. For instance, while *Pokémon1-Same-Positive* and *Pokémon4-Opposite-Negative* both elicited positive ratings, *Pokémon2-Opposite-Positive* and *Pokémon3-Same-Negative* elicited negative ratings. Similarly, when an IAT targeting the two coordination relations was administered participants displayed a clear response bias for *Pokémon1-Same-Positive* relative to

Pokémon3-Same-Negative. The direction of this effect was completely reversed when the two opposition relations were assessed, such that participants displayed a bias for *Pokémon4-Opposite-Negative* relative to *Pokémon2-Opposite-Positive*.

Experiment 2 (Chapter 3) shifted the focus from directly trained relations to those that were entirely derived in nature. Two, three-member coordination relations were engineered consisting of arbitrary Pokémon characters (*Pokémon1-Same-Pokémon2-Same-Pokémon3* and *Pokémon4-Same-Pokémon5-Pokémon6*). Thereafter, an opposition relation was established between Pokémon 1 and positive images and Pokémon 4 and negative images (i.e., *Pokémon1-Opposite-Negative* and *Pokémon4-Opposite-Positive*). Similar to before, evaluative responding was under arbitrarily applicable contextual control such that participants not only evaluated Pokémon 1 positively but also liked Pokémon 2 and 3 as well. Equally, they not only responded negatively to Pokémon 4 but also disliked Pokémon 5 and 6. Comparable outcomes were also obtained when an IAT was administered, with a clear response bias favouring Pokémon 3 relative to Pokémon 6.

In order to increase the generalisability of our findings and protect against method-specific artifacts associated with any one task, Experiment 3 (Chapter 3) replicated the above study using an alternative indirect procedure (IRAP) in place of the IAT. The non-relativistic nature of the IRAP also enabled an investigation of whether brief and immediate relational responding was driven by coordination relations (*Pokémon3-Same-Good* and *Pokémon6-Same-Bad*), opposition relations (*Pokémon3-Opposite-Bad* and *Pokémon6-Opposite-Good*) or some combination of the two. Once again, the three stimuli in the first relation received positive ratings while their counterparts in the second relation were scored negatively. While IRAP performances revealed evidence for derived stimulus relating in-line with prior training, these effects were primarily driven by coordination rather than opposition relations.

Experiment 4 (Chapter 4) further expanded the scope of our relational account in two notable ways. First, fine-grained contextual control was exerted over derived relations consisting of entirely novel stimuli (fictitious brand names and valenced adjectives) and these relations were measured using an IAT as well as a novel indirect procedure (affective priming). Specifically, and following the formation of two coordination relations (*Pardal-Same-Zatte-Same-Ettalas*; *Ciney-Same-Witkap-Same-Gageleer*) half of the participants were trained to relate *Pardal-Same-Positive* and *Ciney-Same-Negative* while the other half were trained to relate *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative*. Consistent with our previous findings, evaluative responding was contingent upon the cue governing the stimulus relation. When a coordination relation was established between Pardal and positive words (or Ciney and negative words) participants rated Pardal, Zatte and Ettalas positively and Ciney, Witkap and Gageleer negatively. However, when an opposition relation was formed between Pardal and positive words (or Ciney and negative words) this preferential pattern was completely reversed; with Pardal, Zatte and Ettalas eliciting negative and Ciney, Witkap and Gageleer positive ratings. Broadly comparable effects were also obtained when an IAT or affective priming task was used to assess responding towards the final brand names in both relations (i.e., Ettalas and Gageleer). Second, derived stimulus relating not only predicted direct and indirect task performance but consumer choice behaviours as well. Once the various measures of evaluation were completed, participants were presented with samples of the six brands and offered the opportunity to take three as a reward. A majority of participants chose Pardal, Zatte and Ettalas when *Pardal-Same-Positive* and *Ciney-Same-Negative*. However, relating *Pardal-Opposite-Positive* and *Ciney-Opposite-Negative* resulted in a different set of brands being selected (i.e., Ciney, Witkap and Gageleer).

Experiment 5 (Chapter 5) revealed that an entirely different pattern of evaluative responding emerges when stimuli participate in comparative rather than coordination or

opposition relations. Two contextual cues meaning ‘More than’ and ‘Less than’ were generated and used to form a five member comparative relation between fictitious prizes (*Pardal < Zatte < Ettalas < Ciney < Witkap*). A consequential function was then established for the first two prizes in the relation by making access to different quantities of money contingent on their selection. As predicted, self-report ratings varied according to a stimulus’ location within the comparative relation, with Ciney evaluated far more positively than either Ettalas or Zatte, and Ettalas evaluated as less positive than Ciney but more so than Zatte. Evidence for comparative relating was also obtained on both the IAT and behavioural choice tasks, with participants favouring Ettalas relative to Zatte, Ciney relative to Ettalas and Ciney relative to Zatte on the former and Ciney on the latter.

In order to further increase the generalisability of our findings and ensure that exposure to a derivation test did not influence evaluative responding, Experiment 6 (Chapter 5) replicated the above study while controlling for a number of method-related factors. Once again, the three stimuli were comparatively related to one another such that the first stimulus was liked less than the second and the second less than the third on both direct and indirect procedures. Participants also consistently chose to take the prize furthest along the comparative relation when a behavioural choice task was administered.

Finally, and perhaps most importantly, the foregoing patterns of evaluative responding only emerged when participants successfully derived the relation between the trained stimuli. In each and every case where a test for derivation was included (Experiments 2-6) a clear, consistent and predictable trend was evident. Participants who passed the derivation test responded in a consistent and systematic manner to the various Pokémon characters, brand products or prizes in-line with the derived relations established among those stimuli. In direct contrast, those who failed the derivation test responded ambivalently towards the stimuli they encountered and often showed incoherent - if any - outcomes on

indirect and behavioural choice tasks. Having summarized the current research programme, the following sections will consider some of the conceptual and theoretical issues raised by the above findings.

6.4 Conceptual Issues

The present studies have important implications for learning theorists and psychologists interested in the study of human likes and dislikes. First and foremost, Relational Frame Theory argues that once a protracted history of arbitrarily applicable relational responding is in place, everything that an individual does thereafter involves verbal processes. Indeed, it is difficult to conceptualize a behavior that is not impacted in some way, shape or form by a prior history of relational learning. One may argue that “non-verbal” responding is *in principle* possible but pragmatically speaking there appears to be no widely agreed upon way to demonstrate a genuinely “non-verbal” act in a verbally sophisticated human.

To illustrate this point more clearly, consider a parrot that has received a direct history of reinforcement for emitting the echoic response “Pretty Polly.” Now consider a verbally-able human who is “trained” to emit this echoic response hundreds of times. After such training both parrot and human will respond with a high degree of speed and accuracy when prompted appropriately. Imagine, however, that both parrot and human are then asked “Who’s opposite to Pretty Polly?”. The parrot will likely respond immediately with the highly trained echoic (“Pretty Polly”) whereas the human may respond appropriately with “Ugly Polly,” thus showing that in the latter case the response participates in a network of arbitrarily applicable relational responses that extends beyond the directly trained echoic. Even in a case where the human simply repeats “Pretty Polly” this does not necessarily render the response non-verbal. For example, imagine the person does respond with “Ugly Polly” when subsequently prompted with “No, no, I said, who is *opposite* to ‘Pretty Polly’.” In other

words, directly trained responses cannot be defined as “non-verbal” on an *a priori* basis for an individual with a verbal history. Instead, for the verbal human, it seems more accurate to say that the echoic response has been derived many times rather than being “non-verbal.” Like Adam and Eve, once we have bitten from the apple of “verbal behavior” there is no going back to a non-verbal Garden of Eden.

If correct, then RFT has important implications for the study of evaluative responding in general and evaluative conditioning in particular. Imagine, for example, that an adult is asked to complete a respondent learning task where a non-sense word (CS) is repeatedly paired with an electric shock (US) across one hundred separate trials. Following training, the previously neutral CS will likely acquire an aversive function, and by implication, elicit negative responses on a stimulus rating task. Defining this change in behavior due to the pairing of stimuli as “non-verbal” or as an instance of “direct contingency learning” assumes that the individual’s history of derived relational responding did not “mediate” that effect but was somehow “switched off” or set aside during the task. This assumption is exceptionally problematic from a functional perspective given that learning history is a core causal concept that is used to understand, predict and influence behavior. Thus while changes in liking due to the pairing of stimuli may be respondent at the procedural level we would argue that they reflect an instance of arbitrarily applicable relational responding at the (behavioral) process level – at least when observed with verbally-able humans. According to this perspective, EC effects based on spatio-temporal contiguity are not a simple form of (respondent) learning as often claimed but rather the most rudimentary form of a complex human behavior learned early on in our development (i.e., derived relational responding).

Although this “EC as relational responding” argument has received primarily empirical support (e.g., Smyth et al., 2006) a more detailed experimental analysis is clearly needed. With this in mind, future work could test the above assumption by demonstrating the

emergence of derived relational performances when stimuli are simply paired with one another in space and time. Moreover, by employing human populations with and without a protracted history of relational learning, as well as a number of other species such as bonobos and chimpanzees (Rumbaugh, Savage-Rumbaugh, King, & Taglialatela, 2011), sea-lions (Kastak, Schusterman & Kastak, 2001) and pigeons (Urcuioli, 2008), researchers could determine whether these derived (evaluative) performances are unique to humans or evident elsewhere in the animal kingdom. On a related note, RFT may also shed light on a number of ambiguous or contradictory findings within the EC literature itself, such as why EC effects are often smaller in populations with a relatively short history of arbitrarily applicable relational responding (children) compared to those with an extended history of relating (adults); why EC effects are larger for novel stimuli without a prior history of relating relative to those with pre-existing stimulus functions; why EC effects occur even when the related stimuli share no formal or physical properties and why those effects are often resistant to extinction following CS only presentations (see Hofmann et al., 2010).

In short and when combined with previous work from the RFT literature, the above findings call for a reinterpretation of evaluative responding as a contextually controlled behavior rooted in a prior history of arbitrarily applicable relational responding. Whereas non-human animals seem shackled to approaching or avoiding stimuli on the basis of direct-contingency learning, humans can spontaneously come to like and dislike stimuli once they participate in derived relations with other valenced stimuli. These evaluative responses can vary in their complexity, stretching from simple mutually entailed relations to more complex (combinatorially entailed) relations all the way up to the relating of relational networks with one another. Although relational responses that are low in complexity and have been repeatedly derived may be topographically similar to behaviors that have been established for non-human organisms via direct contingencies, this does not necessarily mean that they are

functionally similar. Indeed, while it may be tempting to view behavior as verbal when it involves derived relational responding and non-verbal when it involves contact with direct contingencies of reinforcement or stimulus pairing, we argue that this distinction may be problematic for several reasons (for a more detailed treatment see Hughes et al., in press).

A second implication of the current work is that it may stimulate novel theoretical debate given that our findings extend beyond the explanatory scope of associative learning accounts. For several decades now, EC models based on respondent processes have guided a sizeable amount of empirical, methodological and theoretical output, to the point that many social, cognitive, consumer, health, and clinical phenomena are framed as the consequence of stimulus pairings. Given the exceptional success of these models EC researchers might also be tempted to interpret the above effects in terms of respondent learning as well. We believe that this analytic strategy may be problematic for one simple reason: associative learning requires that a direct contingency between the CS and US has been encountered by the individual at some point in time. In the current thesis, however, a direct contingency was *never* established between the first and final stimuli in a derived relation. For example, we never paired Pokémon 1 and 3 or 4 and 6 (Experiments 2-3), Pardal and Ettalas or Ciney and Gageleer (Experiment 4) or even Zatte and Ciney (Experiments 5-6). Nevertheless, participants still responded positively and negatively towards those various stimuli. What about an explanation based on stimulus generalization? This also seems problematic given that generalization requires a prior history of learning with stimuli that are formally similar to the object being generalized to. In the current research, none of Pokémon, brand products or prizes shared any consistent physical properties and their allocation within and between relations was often randomized across participants.

Researchers might still be tempted to explain the observed effects in terms of some form of higher-order classical conditioning. Once again, this seems problematic given that we

used a unidirectional training procedure similar to Smyth and colleagues (2006). In other words, we established a linear chain of stimulus relations such that Stimulus 1 was related to Stimulus 2 and Stimulus 2 was subsequently related to Stimulus 3. Following this training, an evaluative function was then established for Stimulus 1. A transformation of functions from Stimulus 1 to 3 cannot occur through forward Pavlovian conditioning given that Stimulus 3 followed, rather than preceded the presentation of Stimulus 1 and 2 (i.e., forward Pavlovian conditioning requires that the CS predicts the subsequent occurrence of the US). Likewise, an explanation in terms of backward Pavlovian conditioning is equally unlikely given that Stimulus 3 was always separated from the US by two intervening stimuli (i.e., backward Pavlovian conditioning requires that a CS immediately follows the US). Finally, a second-order conditioning account in which a CS1 is first paired with a US, followed by the pairing of CS1 with CS2 also appears to be unsatisfactory given that the evaluative functions of Stimulus 1 were only established *after* the relation was formed - not before. Once Stimulus 1 had been paired with valenced stimuli participants immediately progressed to the direct and indirect measures of evaluation, and as such, no further training was provided in which that stimulus could be related to Stimulus 2 or 3. Indeed, even an account that appeals to some complex combination of these factors (e.g., backward higher-order sensory pre-conditioning) faces the additional difficulty of explaining why the trained functions for Stimulus 1 were opposite to (Experiments 2-4) or less than (Experiments 5-6) the untrained functions acquired by Stimulus 3. When taken together, the viability of higher-order conditioning as a potential explanation of our transformation effects seems rather unlikely.

Although respondent models of learning fail to accommodate the effects observed above, these findings do fit well with RFT's claim of contextually controlled relational responding as a generalized operant acquired through abstraction or multiple exemplar training. In Experiments 1-6, operant contingencies were applied across multiple sets of

stimuli and these contingencies successfully established the evaluative responses observed on direct and indirect procedures alike. When taken together, our findings affirm the central role that relational learning plays in the formation of evaluative responding and the need to explore behavioural processes and preparations of a non-respondent nature. More generally, RFT appears to represent a theory capable of accommodating the interaction between direct and derived stimulus relating and may offer a more complete account of evaluative responding than direct-contingency accounts alone.

Third, our findings support and extend the notion that derived stimulus relations can be generated using procedures above and beyond the Matching-To-Sample (MTS) protocol. Until recently, the majority of transformation studies in the RFT literature have relied on some variant of MTS to train and test laboratory induced relations (e.g., Dougher et al., 2007; Gil et al., 2012; Whelan & Barnes-Homes, 2004). This over-reliance on a single methodology may not only undermine a flexible and progressive analysis of derived relational responding but also introduce concerns that these performances are due to some specific property of MTS itself (see Dymond & Whelan, 2010). In order to circumvent these issues, alternative procedures have recently been introduced to train and/or test for derived relational responding. Examples include stimulus pairing tasks (Leader & Barnes-Holmes, 2001), Go/No-go tasks (e.g., Cullinan, Barnes-Holmes, & Smeets, 2001) the Relational Evaluation Procedure (Stewart, Barnes-Holmes & Roche, 2004) and the Relational Completion Procedure (Dymond & Whelan, 2010). The above set of studies is the first to show that derived relations may be successfully formed when a modified operant version of the Picture-Picture paradigm is used. On a related note, only a single transformation study has employed the IAT to date (O'Toole et al., 2007) and the authors focused exclusively on equivalence relations. As such, this is the first time that a derived transformation of evaluative functions through coordination, opposition and comparative relations has been shown to produce

differential outcomes on the IAT, IRAP, affective priming and direct measurement procedures.

Finally, the current thesis set out with two complimentary goals in mind. The first was to apply RFT to the study of evaluative responding, and in doing so, show that better prediction-and-influence of behaviour can be achieved when the role of derived stimulus relating is explicitly acknowledged. At the same time we adopted a wide variety of direct and indirect tasks in order to put our second goal to the test (i.e., to demonstrate that “implicit” or “automatic” evaluative responding is inherently relational in nature). Similar to evaluative conditioning, the implicit cognition literature has historically been dominated by mechanistic (mental) models that frame “automatic” and “controlled” behaviour as the product of either single, dual or multiple mental processes operating exclusively or in interaction (for a review see Gawronski & Sritharan, 2010). More often than not, these various accounts make the additional assumption that automatic behaviours are mediated by the formation, activation and change of unqualified associations in memory.

Unsurprisingly, a number of contextual behavioural scientists have rejected the notion of mediating mental forces/agency and have sought to accommodate the outcomes observed on direct and indirect tasks in terms of causal relations between the environment and behaviour. One functional account in particular, known as the Relational Elaboration and Coherence (REC) model, argues that the behaviour obtained on a measurement procedure reflects an interaction between the individual’s learning history with respect to the targeted relation(s) and the specific features of the context in which they are assessed (Hughes et al., in press). More precisely, the REC model suggests that arbitrarily applicable relational responding can be carved into different patterns of behaviour that vary in their relative levels of relational complexity and derivation. Behaviours that involve *relatively* lower levels of time, complexity and derivation are often labelled as brief and immediate relational responses

(BIRRs) while those that involve relatively higher levels of these respective response properties are labelled extended and elaborated relational responses (EERRs).

Although a growing number of studies support the notion that implicit cognition is relational rather than strictly associative in nature, the bulk of this work has focused on pre-experimentally established relational repertoires (see Hughes & Barnes-Holmes, in press). Strong support for the REC model - at least from a functional perspective - awaits a systematic exploration of the learning histories and current contextual variables necessary to establish, maintain, and change BIRRs and EERRs within the laboratory. We believe that the current work represents an important first step in this general direction. Experiments 1-6 clearly identify the learning history necessary to generate a wide variety of relational outcomes on tasks such as the IAT, IRAP and affective priming. Note, however, that while we manipulated relational complexity and derivation (and demonstrated that low complexity relations that are derived within a single session of learning can drive direct and indirect effects) we failed to do so in a systematic fashion. For example, while mutually entailed relations were generated in Experiment 1, combinatorially entailed coordination and opposition relations in Experiments 2-4 as well as comparative relations in Experiments 5-6, we did not compare differences in the relative speed and accuracy with which participants responded based on these different relations (see Steele & Hayes, 1991; O'Hora, Roche, Barnes-Holmes & Smeets, 2002 for work in this vein). Likewise, while we altered the relative location of the derivation test in Experiments 2 and 6, we did not deprive any participants of this "context for deriving", nor did we vary the number of opportunities to derive (e.g., 1 versus 100) within and between those tests. Furthermore, we did not attempt to manipulate the degree to which laboratory induced BIRRs and EERRs cohered with one another across different procedures. Consequently, future research will need to provide a more rigorous test of the REC model's core assumptions by subjecting relational complexity

and derivation, as well as their interaction, to a detailed experimental analysis. To our knowledge no research has directly explored the intersection between these two factors in the production of different relational responses (i.e., BIRRs and EERRs) or even their relation to other properties of relational responding that may play an important role in implicit cognition (e.g., relational coherence). Embarking on the above research agendas would serve to refine our understanding of when and why BIRRs and EERRs independently or interactively predict different classes of behaviour.

6.5 Implications for the Mechanistic Approach to Evaluative Responding

Although we have provided an explanation of evaluative responding in terms of functional interactions between the environment and behaviour, the current data may also be of interest to researchers operating at the mental (mechanistic) level of analysis. As outlined in Chapter 2, three mental models currently dominate discourse within this research area and accommodate our findings with differing levels of success. We will now consider each of these models in turn.

Associative Mental Models

The above results appear to seriously conflict with associative accounts of evaluative responding (Baeyens et al., 1992; Martin & Levey, 1994). According to this perspective, the direct pairing of stimuli leads to the formation of unqualified links between mental representations in memory. Thus presenting one stimulus with positive images and another with negative images should result in those stimuli being evaluated in-line with the valence implied by the images they were paired with – regardless of the relation established between those stimuli by a contextual cue. In Experiments 1-3, for example, relating *Pokémon-Same-Positive* and *Pokémon-Opposite-Positive* should have resulted in identical (positive) outcomes on the various measures of evaluation. Similarly, presenting Pardal with positive or Ciney with negative words and reinforcing the selection of the ‘Opposite’ cue should have

resulted in the former brand being rated positively and the latter negatively (Experiment 4). None of these predictions were confirmed in any of the above studies, with the direction - and in the case of Experiments 5 and 6 - magnitude of evaluative responding dictated by the cue governing the stimulus relation.

Associative accounts also face the additional challenge of explaining how stimuli that were never paired with valenced words or images came to elicit responses on direct and indirect procedures. In Experiments 2-3, for instance, Pokémon 3 and 6 should not have elicited evaluative responses given that neither stimulus was directly paired with valenced stimuli nor were they paired with Pokémon 1 or 4 during the task. Likewise, the brands Ettalas and Gageleer in Experiment 4 and the prize Ciney in Experiments 5 and 6 should not have been liked or disliked for the very same reason. Even allowing for some combination of higher-order backward sensory preconditioning (*see above*) these models fail to explain why the psychological functions established for one stimulus were not simply transferred but rather transformed through the derived relations. For instance, when Pokémon 1 was related with negative images using the ‘Opposite’ cue participants *liked* Pokémon 2 and 3 (Experiment 3); when Pardal was *Opposite-Positive* both Zatte and Ettalas were *disliked* (Experiment 4) while Ciney was liked *more* than Ettalas or Zatte in Experiments 5-6. Finally, the fact that indirect effects only emerged when participants correctly derived the relation between stimuli seems to provide the strictest challenge for this class of models. Given that implicit or automatic evaluations are often assumed to be governed by associations in memory formed on the basis of experienced pairings (Gawronski, Deutsch & Banse, 2011; Nosek et al., 2011), IAT, IRAP and affective priming effects should not have been observed at any point in the current thesis. Contrary to this claim, participants displayed evidence of derived coordination, opposition and comparative relating across a variety of indirect

procedures and stimulus sets. As such, traditional associative explanations appear to be problematic on multiple fronts.

Propositional Models

The current data not only undermine associative models but provide firm support for propositional accounts that involve qualified links between mental representations in memory. From this perspective, changes in evaluative responding are moderated by direct and derived stimulus relating which is in turn 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 (e.g., “*X is opposite to Y*” or “*X is larger than Y*”). Likewise, while associations gradually develop with many experienced pairings (Smith & DeCoster, 2000; Strack & Deutsch, 2004), propositions can be formed on the basis of direct training or inferred via deductive reasoning and language (De Houwer, 2009a). When combined these two properties of propositions may explain the current set of findings.

In our studies participants were exposed to contextual cue training in which the selection of an arbitrary symbol was reinforced in the presence of pictures bearing a relation of similarity or opposition. Across numerous different exemplars they abstracted out the relational functions of ‘Same’ for one cue and ‘Opposite’ for the other. From a mechanistic point of view, this history of relational learning could be interpreted as one that gave rise to ‘Same’ and ‘Opposite’ propositions about the two arbitrary symbols (e.g., “*This symbol means that the pictures are the same and that symbol means they are opposite*”). Thereafter, a series of stimulus relations were formed by differentially reinforcing the selection of one of these symbols in the presence of a stimulus combination. Mechanistically speaking, the differential selection of the ‘Same’ and ‘Opposite’ cues across stimulus pairs may have resulted in the formation of additional propositions about the trained relations (e.g.,

“Pokémon 1 is the same as Pokémon 2”, “Pokémon 2 is the same as Pokémon 3” and *“Pokémon 1 is the opposite of negative images”*). Critically, participants who passed the derivation test may have also made a “propositional leap” insofar as they generated a set of novel, untrained propositions about the derived relations (e.g., *“Pokémon 3 is good and Pokémon 6 is bad”*). If this latter assumption is correct, then it was these inferred propositions (rather than their directly trained counterparts) that mediated responding on the direct and indirect procedures. This may help to explain why only participants who successfully derived (i.e., made a “propositional leap”) showed evidence for evaluative responding while their counterparts who failed that test (but passed all stages of contextual and relational training) did not. In other words, a series of propositions based on direct stimulus relating may not have been enough – rather these propositions may have to give rise to additional propositions in order to produce the expected effects.

Proponents of propositional models often argue that the formation of these mental constructs requires not only an awareness of the to-be-related stimuli but the cognitive resources, time and goals to relate them (De Houwer, 2009a). Although we did not manipulate these various mental conditions in the current thesis (given our functional perspective) it could be argued that they were nevertheless present during contextual cue and relational training. For instance, progression through training and testing trials was free from time constraint, required 100% accuracy and did not involve a distractor or non-relevant task during learning (i.e., training was non-time pressured, goal directed, and explicitly required an awareness of the stimuli). While admittedly speculative, mechanistic researchers could replicate the current work while controlling for the above factors. Our findings also introduce the possibility that once formed propositions may be activated automatically from memory and guide the evaluation of stimuli on direct and indirect measures alike (see Hughes, Barnes-Holmes & De Houwer, 2011). For example, once the proposition *“Ciney is more than*

Pardal” has been generated based on the derived relation between those stimuli, this proposition may have been stored in memory. In contrast to dual-process accounts (*see below*) this memory representation is assumed to be propositional in nature. Although mechanistic researchers have often appealed to mental associations in explaining indirect task performances, there is no *a priori* reason why propositional knowledge could not be activated from memory and lead to automatic evaluations of a given stimulus.

Dual-Process Models

Finally, dual-process accounts that allow for rules (Smith & DeCoster, 2000), judgments (Kahneman, 2003), and propositions (Gawronski & Bodenhausen, 2011) to feed into and create novel associations could account for our findings – provided that certain pre-conditions are met. For instance, and similar to the above propositional model, contextual cue training may be mediated by the formation of propositions concerning the meaning of the two symbols. These propositions could then be used during relational training to generate additional propositions about the stimulus relations themselves. This set of “directly trained” propositions could subsequently give rise to novel, untrained propositions about stimuli that were never paired together. If correct then these inferred propositions could be transformed into and stored as associations in memory, explaining the direct and indirect outcomes observed in Experiments 1-6.

On balance, while interactive dual-process models may accommodate our transformation of function effects they do not appear to do so in an *a priori* fashion. Rather it appears that this particular class of models can account for virtually every empirical finding post-hoc. In addition, they often fail to provide clear testable predictions for their falsification or even offer a means to empirically distinguish between the explanatory concepts of one dual-process model relative to another (for a discussion see Gawronski & Creighton, in press; Keren & Schul, 2009). Our findings thus pose a challenge to dual-process (and propositional)

model theorists to articulate how, when and why (a) the contextual cue and relational training procedures give rise to propositions (b) novel propositions will emerge based on a set of directly trained propositions, and in the case of dual-process accounts, (c) how these propositions are subsequently transformed into and stored as associations given that participants were never exposed to any training or instructions to do so. More generally, while the encoding of affirmation or negation propositions seems to fit with traditional notions of mental associations (e.g., “*Pokémon-Good*”) it is not immediately clear how other propositions, such as those involving comparison, distinction, or hierarchy are stored as unqualified links in memory given their inherently relational content.

Summary. In short, mechanistic researchers may explain the current results as being *moderated* by derived stimulus relations encountered in the environment and *mediated* by the formation of either mental associations, propositions or some combination of the two. Within this world-view theoretical models are usually evaluated on two fronts - their ability to account for existing knowledge about the phenomenon of interest (heuristic value) and ability to make *a priori* predictions rather than post-hoc rationalizations (predictive value). Given that evaluative responding only emerged when participants correctly derived the relation between stimuli, and that derivation can conceptually be achieved on the basis of propositions but not associations, the current results provide strong support for propositional (and to a lesser extent dual-process) models and against their purely associative counterparts. It must be said, however, that single process models involving propositions seem to provide a coherent and parsimonious explanation for the above data without the need for recourse to a second (associative) mental construct.

6.6 Interplay Between the Functional and Mechanistic Traditions

In walking the tight-rope between a mechanistic and functional approach to evaluative responding, the current thesis is exquisitely sensitive to the dangers of conflating one with the

other. Although researchers from both traditions are interested in the same behavioural outcomes they differ dramatically in their scientific goals, truth criteria and explanatory concepts. On the surface these philosophical points of departure may seem irreconcilable and suggest that fruitful dialogue is improbable or even undesirable. We believe that such a position is detrimental for all concerned. Instead of subjecting one another to derision and scorn both traditions may profit from the recently proposed functional-cognitive framework advanced by De Houwer (2011). At its core, this account argues that the functional and mechanistic approaches may indeed operate at fundamentally distinct levels of analysis, but they can be mutually supportive, insofar as theoretical, methodological and empirical developments at one level can lead to advances at the other level. The current thesis may represent the first practical application of this collaborative framework in the domain of evaluative responding. In particular, our functional contextual approach - while sufficient in and of itself - may facilitate progression at the mechanistic level in three distinct ways.

Conceptual Implications

First, the above work draws attention to an exciting new behavioural phenomenon that has yet to be systematically explored by mechanistic researchers. Derived stimulus relating represents a purely functional explanation of how, when and why the psychological properties of a stimulus can be altered in the absence of training, reinforcement or instruction. For several decades now, CBS has invested considerable time and effort in exploring the origins of this process, its boundary conditions and interaction with other types of learning. The result is a wealth of research implicating derived relating in the development of perspective taking and self (McHugh & Stewart, 2012), intelligence (Cassidy et al., 2011) and language (Hayes et al., 2001), as well as the etiology of fears, phobias, anxiety and avoidance (Hayes, Strosahl & Wilson, 1999). Likewise, implicit and explicit cognition (Hughes et al., in press), problem-solving (Stewart et al., in press), and analogical reasoning

(Lipkens & Hayes, 2009) all submit to an experimental analysis from this perspective. Despite these remarkable developments, this behavioural process remains an undiscovered country as far as mechanistic research and theorizing is concerned. We believe that this behavioural phenomenon could not only lead to a better prediction and organization of existing findings within mechanistic research but also give rise to new hypotheses and theories about the mental constructs that mediate this process. For instance, derived relating may provide new insight into cognitive consistency (Gawronski & Strack, 2012), expectancy models of avoidance learning (Dymond et al., 2011), the origins of false-memories (Guinther & Dougher, 2010), theory of mind and development of deception (McHugh et al., 2007), not to mention attitude research (Hughes et al., 2011), altruism (Vilardaga & Hayes, in press), transitive inferences (Munnely, Dymond & Hinton, 2010) and behavioural/cultural change (Wilson, Hayes, Biglan & Emby, in press).

Empirical Implications

Second, and more generally, the functional literature represents a vast untapped reservoir of behavioural effects that mechanistic researchers can use to (a) better identify the environmental determinants of a given behaviour and (b) improve existing or develop new mental theories. To illustrate, consider the contextually-controlled patterns of evaluative responding observed throughout the current thesis. Defining these effects in a purely functional manner provides valuable information about the interplay between environment and behaviour. For instance, our results show that reinforcing the selection of a laboratory trained contextual cue rather than some other factor (e.g., mere pairing of a CS and US) results in changes in responding. Similarly, we now know that establishing a series of stimulus relations through differential reinforcement gives rise to a number of untrained derived relations between those stimuli. Note that the functional knowledge acquired in Experiments 1-6 is abstract in nature. Rather than a simple compilation of behaviour effects,

our findings allow us to make inferences about a generalized learning process that applies inside and outside of the laboratory. In other words the current data can be abstracted in order to understand why certain evaluative responses have occurred in the past not to mention predict and influence their probability in the future. This is all achieved while maintaining a firm separation between the concepts used to explain (i.e., environment) and that which needs to be explained (i.e., behaviour) (see De Houwer, 2011a).

In our view, this rapidly growing body of functional knowledge could trigger a new wave of theoretical innovation at the mechanistic level of analysis. Given that the functional approach is solely focused on the interplay between the environment and behaviour it does not place *a priori* restrictions on what mental processes or representations mediate those functional interactions. Thus with respect to the current work, researchers can deploy associations and/or propositions (De Houwer, 2009a; Gawronski & Bodenhausen, 2011; Olson & Fazio, 2009), impressions and judgments (Kahneman, 2003), reflexive-impulsive systems (Strack & Deutsch, 2004) or any other mechanistic concepts to explain how the environment influenced the observed outcomes. Critically, however, behavioural effects from the functional literature do constrain and refine mental theorizing in an *a posteriori* fashion. For instance, the finding that evaluative responding only emerges when participants show evidence of derivation may cause associative explanations to be discarded and those involving propositions to be selected out. Furthermore, evidence of coordination, opposition and comparative relating across a number of indirect procedures could also hold important implications for mental accounts of implicit cognition. In short, the functional literature represents a goldmine of empirical findings that specify the reciprocal relationship between environment and behaviour. For mechanistic researchers these effects could provide another means to select between competing theoretical accounts without imposing any limitations on the emergence of those theories in the first instance. Indeed, a more sophisticated

appreciation of the environmental moderators of behaviour can only help (rather than hinder) developments at the mechanistic level of analysis.

Methodological Implications

Third and finally, the current work may also equip mechanistic researchers with a number of novel methodologies for generating direct and derived stimulus relations. These procedures could then be used to explain how mental constructs such as propositions potentially mediate the above relational performances. For example, our contextual cue and relational training procedures may represent a method for generating different types of propositions that vary in their respective complexity. Likewise, the derivation test may represent a useful context within which participants can evaluate the validity of recently formed propositions. When combined these procedures may allow for the formation of propositions to be manipulated independently from their subsequent truth evaluation (see De Houwer, 2009a). At the same time, and as noted in Experiment 3, the IRAP represents the only indirect procedure in the literature that can target different types of brief and immediate relational responses in a non-relativistic fashion. Unlike the IAT, AMP or affective priming, the IRAP can identify whether participants respond with greater speed and accuracy when relating *Slim People-Good-Similar* independently of *Fat People-Bad-Opposite* (Roddy, Stewart & Barnes-Holmes, 2011), *White People-Safe-True* from *Black People-Dangerous-False* (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010) and *Adult-Sexual-False* from *Child-Non-Sexual-True* (Dawson, Barnes-Holmes, Gresswell, Hart & Gore, 2009). In addition, the IRAP can also identify what brief and immediate relational responses are most predictive of meaningful “real-world” behaviours, such as cocaine users adherence to an inpatient treatment program (Carpenter, Martinez, Vadhan, Barnes-Holmes & Nunes, 2012) or the likelihood of high-spider fearful participants approaching a live tarantula (Nicholson & Barnes-Holmes, 2012). Consequently, this novel indirect procedure may afford mechanistic

researchers with a means to test the automaticity of recently formed propositions in a manner that is beyond the capacity of other indirect procedures within the research literature.

Note that the benefits of the functional-cognitive framework are by no means unidirectional. The mechanistic approach may aid contextual behavioural scientists in refining their functional theories through access to novel procedures and research areas. Our work provides a ready demonstration of this point: we developed the contextual cue and relational training procedures by modifying a well-known social/cognitive task (the Picture-Picture paradigm) and imported indirect methods from this literature to test our relational account. More generally, the emergence of implicit cognition as a distinct subject of inquiry has spurred a flurry of theoretical and methodological development at the functional level – most notably in the form of the IRAP and REC Model (see Hughes et al., in press). Thus, mechanistic researchers may introduce their functional counterparts to previously undiscovered intellectual domains while equipping them with useful tools that they can use to identify new functional relations between the environment and behaviour.

Summary. It should now be apparent that the functional and mechanistic levels of analysis are mutually supportive so long as a clear separation is maintained between their core assumptions, values and scientific goals. Increasing our ability to understand, predict and influence the environmental determinants that moderate evaluative responding only serves to strengthen theories concerning the mental mediators of evaluation. Likewise, debate surrounding mental mechanisms and their operating conditions can shine a light on exciting new behavioural domains and spur continued methodological innovation at the functional level. Although we have confined our application of this framework to the study of evaluative responding, it seems plausible that its wider adoption throughout psychological science may accelerate empirical, methodological and theoretical progress in other domains as well.

6.7 Limitations and Future Directions

Although the current research programme is consistent with our relational account, a number of questions still require attention. First, we found that evaluative functions were transferred through coordination, reversed in opposition and lead to relativistic differences when comparative relations were involved. It remains to be seen whether other relations also have a signature “fingerprint” when it comes to the transformation of functions from one stimulus to another. Imagine, for example, that four mutually entailed causal relations were formed between arbitrary stimuli and appetitive or aversive events using cues meaning “Causes” and “Prevents” (e.g., *A-Causes-Money*; *B-Prevents-Money*; *C-Causes-Cancer* and *D-Prevents-Cancer*). In this instance broadly similar outcomes may be obtained as in coordination and opposition relations, such that participants respond positively to A and D and negatively to B and C (see Moran & Bar-Anan, in press). However, when these causal relations involve combinatorially entailed stimuli, a comparative pattern of evaluative responding may emerge. For instance, training a three member causal relation (e.g., *Sonin-Causes-Xanthan-Causes-Pixin*) and then informing participants that Pixin is another word for Cancer may result in Pixin being rated more negatively than Xanthan and Xanthan more negatively than Sonin (given that Sonin indirectly causes Xanthan while Xanthan directly causes Cancer; see Waldmann & Hagmayer, 2005 for related findings).

Future research could therefore expand the scope of our relational account by exploring whether causal, temporal (“Before/After”), deictic (“I/You”) and other stimulus relations give rise to distinct patterns of evaluative responding. This work could also model more complex transformations of function by generating a number of different stimulus relations and then relating them hierarchically with one another (see Gil et al., 2012). Doing so would demonstrate that the same stimulus could elicit both positive and negative responses depending on the specific relation being assessed at any given point in time.

Second, we limited our analysis to the formation of derived relations and remained comparatively silent to their subsequent modification. In future work, it will be critical to investigate whether (a) altering the “structure” of an experimentally established relation or (b) the psychological functions of stimuli within that relation leads to immediate changes in the functions of other related stimuli. Consider the first strategy: altering the manner in which stimuli are related to one another. A three-member coordination relation could be generated (*A-Same-B-Same-C*) and a respondent function established for the A stimulus by pairing it with shock. Upon testing we would expect that A, B and C would each elicit broadly similar fear responses. However, if participants were subsequently re-trained to relate *A-Opposite-B-Same-C* a different pattern of responding should emerge. That is, A should still elicit fear responding whereas both B and C should acquire appetitive functions. Alternatively, if *A-Same-B-Opposite-C* another pattern of responding should be anticipated and still another if *A-Opposite-B-Less than-C*. In other words, subsequent modifications to the “structure” of a relation may result in a cascade of novel functions being transformed through that relation, and by implication, different patterns of evaluative responding. Note that these changes in responding should occur without any modifications to the original functions of the A stimulus itself. In the above example, for instance, the B and C never predict shock yet they still acquire appetitive functions in one context and aversive functions in another by virtue of their participation in a derived relation with the A stimulus.

Now consider the second modification strategy: extinguishing the functions of a stimulus that acquired those functions through either direct training or derivation. To illustrate this more clearly, imagine that a five-member comparative relation was trained (e.g., $A < B < C < D < E$) and a respondent function was established for the C stimulus by pairing it with either an aversive (loud scream) or appetitive (melodic piece of music) sound. Administering direct and indirect measures of evaluation should lead to similar effects as

seen in Experiments 5 and 6. Researchers could then examine whether those functions are eliminated when either the E stimulus (i.e., derived extinction) or the C stimulus (direct extinction) are repeatedly presented without any auditory stimulus. Drawing on past research we would expect that all of the previously liked and disliked stimuli should elicit largely ambivalent responses despite the fact that only a single stimulus in the relation had its functions extinguished (e.g., Dougher et al., 1994; Roche et al., 2008). Future work could disentangle these possibilities by comparing the influence of direct versus derived extinction on different types of stimulus relations, and by implication, evaluative responding.

Third, evaluative conditioning researchers may question the relevance or applicability of the current findings given that (a) stimuli were not paired with one another but related in the presence of contextual cues and (b) a stimulus-response contingency was employed throughout all sections of training and testing. Critically, however, an increasing number of authors are employing these very same manipulations within the EC literature. Contextual cues such as “Loves/Loathes” (Förderer & Unkelbach, 2011), “Start/Stop”, “Allow/Prevent” (Moran & Bar-Anan, in press) and “Like/Hate” (Walther et al., 2011) have all been used to relate a CS and US (see also Zanon et al., 2012). Others have sought to pair a CS and US indirectly by embedding them within complex verbal narratives (Gregg et al., 2006), statements (De Houwer, 2006), and instructions (Balas & Gawronski, 2012). Still others have employed stimulus response contingencies that arguably involve some combination of operant and respondent learning (e.g., Beckers et al., 2002; Gast & De Houwer, 2012; Jones et al., 2009; Olson & Fazio, 2001). When taken together, this work indicates that there is considerable scope with which to interpret the notion of “changes in liking resulting from the pairing of stimuli”.

Nevertheless, future work could more closely align the EC and RFT literatures by combining the protocols developed in the current thesis with another task known as the

stimulus pairing observation procedure (SPOP). Developed by Leader and Barnes (1996) the SPOP simply requires participants to observe contiguous presentations of stimuli and not demonstrate any overt response. When subsequently tested, participants consistently show evidence of having formed derived relations between those stimuli (Smeets et al., 1997; Leader et al., 2000) as well as having transformed functions through those relations (Smyth et al., 2006). In other words, the SPOP appears to give rise to EC effects that are relational rather than strictly associative in nature. As such it may provide a more convincing case that while EC procedures involve the associative pairing of stimuli, the underlying behavioural process that gives rise to those effects is distinctly non-associative.

One potential limitation of the SPOP - at least as currently conceived - is that it can only generate equivalence relations. However, by interfacing this procedure with contextual cue training more complex relations could be created. For instance, and similar to Experiments 1-4, cue training could be used to generate the relational functions of 'Same' and 'Opposite' for two arbitrary symbols. Thereafter, SPOP training could be provided in which A1 is paired with B1; A2 with B2; B1 with C1 and B2 with C2. The selection of the 'Opposite' cue could then be reinforced in the presence of A1 and an appetitive stimulus or A2 and an aversive stimulus. If derived relations have been formed as predicted, then participants should rate A1, B1 and C1 negatively and A2, B2 and C2 positively. Alternatively, these cues could be buried in the SPOP itself such that A1 and B1 are both presented together with a symbol meaning 'Less than'. If a respondent function is then established for A1 a comparative set of evaluative responses should emerge similar to Experiments 5-6.

6.8 Conclusion

The current thesis represents a deliberate attempt to contribute to both the functional and mechanistic intellectual traditions. On the one hand, we offer CBS researchers an account

of evaluative responding framed in terms functional interactions between environment and behaviour. The central role of derived relational responding is highlighted and the remit of Relational Frame Theory extended to a new behavioural domain (i.e., human likes and dislikes). On the other hand, we provide mechanistic researchers with a coherent package of studies that not only introduce a novel behavioural phenomenon but also facilitate the development of cognitive explanations of mental evaluation. In either case, we propose that a comprehensive understanding of human likes and dislikes requires a shift in current research practices, such that the influence of direct and derived stimulus relations are explored in tandem.

References

- Augustson, E. M., Dougher, M. J., Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record, 50*, 745–770.
- 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. *Behaviour 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.
- Bar-Anan, Y., & Dahan, N. (in press). The effect of comparative context on evaluative conditioning. *Cognition and Emotion*.
- 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.
- Bar-Anan, Y., & Nosek, B. A. (2012). A Comparative Investigation of Seven Implicit Measures of Social Cognition. *Unpublished manuscript*.
- Bar-Anan Y., & Nosek, B. A., & Vianello, M. (2009). The Sorting Paired Features Task: A Measure of Association Strengths. *Experimental Psychology, 56*, 329-343.

- 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, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and non-arbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, *59*, 61-81.
- Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behaviour: Two perspective on the concept of generalized operant classes. *The Psychological Record*, *50*(2), 251-265.
- Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P. M. (2004). Establishing relational responding in accordance with opposite as generalized operant behaviour in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 559-586.
- 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.
- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*, 527-542.
- Barnes-Holmes, D., Hayes, S. C., & Barnes-Holmes, Y. (2012). Derived stimulus relations as learned behaviour: A modern behavioural approach to human language and cognition. *Unpublished manuscript*.
- 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, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure (IRAP): Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57-66.
- Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *The Psychological Record, 60*, 287-306.
- Barnes-Holmes, D., Regan, D., Barnes-Holmes, Y., Commins, S., Walsh, D., Stewart, I., Smeets, P. M., Whelan, R., & Dymond, S. (2005). Relating derived relations as a model of analogical reasoning reaction times and event-related potentials. *Journal of the Experimental Analysis of Behavior, 84*(3), 435-451.
- Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test (IAT): Measuring attitudes towards Dublin and country life in Ireland. *The Psychological Record, 59*, 389-406.
- Bechtel, W. (2008). Mechanisms in cognitive psychology: What are the operations? *Philosophy of Science, 75*, 995-1007.
- Beckers, T., De Houwer, J., & Eelen, P. (2002). Automatic integration of non-perceptual action effect features: The case of the associative affective Simon effect. *Psychological Research, 66*, 166-173.
- Beckers, T., de Vicq, P., & Baeyens, F. (2009). Evaluative conditioning is insensitive to blocking. *Psychologica Belgica, 49*, 41-57.
- Biglan, A., & Hayes, S. C. (1996). Should the behavioural sciences become more pragmatic?

- The case for functional contextualism in research on human behaviour. *Applied and Preventive Psychology: Current Scientific Perspectives*, 5, 47-57.
- Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation*, 39, 58–70.
- 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.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology*, 28, 207–233.
- 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.
- Cassidy, S., Roche, B. & Hayes, S. C. (2011). A relational frame training intervention to raise Intelligence Quotients: A pilot study. *The Psychological Record*, 61, 173-198.
- 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.

- Critchfield, T. (2011). Translational contributions of the experimental analysis of behaviour. *The Behavior Analyst, 34*, 3–17.
- Cullinan, V., Barnes-Holmes, D., & Smeets, P. M. (2001). A precursor to the relational evaluation procedure: The search for the contextual cues that control equivalence responding. *Journal of the Experimental Analysis of Behavior, 76*, 339–349.
- Dack, C., Reed, P., & McHugh, L. (2010). Multiple determinants of transfer of evaluative function after conditioning with free operant schedules of reinforcement. *Learning and Behavior, 38*, 348–366.
- Davey, G. C. L. (1994). Defining the important theoretical questions to ask about evaluative conditioning: A reply to Martin and Levey (1994). *Behaviour Research and Therapy, 32*, 307–310.
- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J. P., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment, 21*, 57–75.
- 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. (2009a). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*, 1–20.
- De Houwer, J. (2009b). Conditioning as a source of liking: There is nothing simple about it. In Wänke, M. (Ed.). *Frontiers of Social Psychology: The Social Psychology of Consumer Behavior*. New York: Psychology Press.

- De Houwer, J. (2011a). 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. (2011b). 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., 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.
- Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on “automatic evaluations” as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, 45, 101-114.
- Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, 46, 243-257.
- 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., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, 62, 331-351.

- 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.
- Dougher, M., Perkins, D. R., Greenway, D., Koons, A., & Chiasson, C. (2002). Contextual control of equivalence-based transformation of functions. *Journal of the Experimental Analysis of Behavior*, 78, 63–93.
- 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., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more-than, and less-than. *Journal of the Experimental Analysis of Behavior*, 64, 163-184.
- Dymond, S., Bateman, H., & Dixon, M. R. (2010). Derived transformation of children's pre-gambling game playing. *Journal of the Experimental Analysis of Behavior*, 94, 353-363.
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behaviour: The transformation of stimulus functions. *The Behavior Analyst*, 23, 239–254.
- 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.
- Dymond, S., & Whelan, R. (2010). Derived relational responding: A comparison of matching-to-sample and the relational completion procedure. *Journal of the Experimental Analysis of Behavior*, 94, 37–55.
- Ebert, I. D., Steffens, M. C., von Stulpnagel, R., & Jelenec, P. (2009). How to like yourself

- better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology*, 45, 1098–1104.
- 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., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes, *Cognition & Emotion*, 25(4), 639-656.
- Field, A. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, 26, 857–875.
- Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A non-associative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 211–224.
- 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.
- Fox, E. J. (2006). Constructing a pragmatic science of learning and instruction with functional contextualism. *Educational Technology Research & Development*, 54(1), 5-36.
- 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 behavioural analysis of human anxiety. *The Psychological Record, 61*, 227-252.
- 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., & 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.
- Gaudio, B. A. (2011). A review of acceptance and commitment therapy (ACT) and recommendations for continued scientific advancement. *The Scientific Review of Mental Health Practice, 8*, 5-22.
- 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., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 78-123). New York, NY: Guilford Press.
- 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., & Strack, F. (Eds.). (2012). *Cognitive consistency: A fundamental principle in social cognition*. New York, NY: Guilford Press.
- 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.
- Gifford, E. V., & Hayes, S. C. (1999). Functional contextualism: A pragmatic philosophy for behavioural science. In W. O'Donohue & R. Kitchener (Eds.), *Handbook of behaviorism* (pp. 285-327). San Diego: Academic Press.
- 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.
- 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.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197-216.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17– 41.
- 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.

- Guinther, P. M., & Dougher, M. J. (2010). Semantic false memories in the form of derived relational intrusions following training. *Journal of the Experimental Analysis of Behavior, 93*, 329–347.
- 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. (2004). Acceptance and Commitment Therapy, Relational Frame Theory, and the third wave of behaviour therapy. *Behavior Therapy, 35*, 639-665.
- 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 Behavioural Science: Creating a Science More Adequate to the Challenge of the Human Condition. *Journal of Contextual Behavioural Science, 1*(1), 1-16.
- Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behaviour-analytic view of the purposes of science. *The Behavior Analyst, 9*(2), 175-190.
- Hayes, S. C., Hayes, L. J., & Reese, H. W. (1988). Finding the philosophical core: A review of Stephen C. Pepper's World Hypotheses. *Journal of the Experimental Analysis of Behavior, 50*, 97–111.

- 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.
- Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2004). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy*, *43*, 533–551.
- 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.
- Holth, P., & Arntzen, E. (1998). Stimulus Familiarity and the Delayed Emergence of Stimulus Equivalence or Consistent Nonequivalence. *The Psychological Record*, *48*(1), 81-110.
- 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 of alcohol-related attitudes craving and behaviour. *Addictive Behaviours*, *35*(12), 1161-1163.
- Hughes, S., & Barnes-Holmes, D. (in press). A Functional Approach to the Study of Implicit Cognition: The Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. To appear in Dymond, S & Roche, B. (Eds.). *Advances in Relational Frame Theory & Contextual Behavioural Science: Research & Application*. Oakland, CA: New Harbinger Publications.
- Hughes, S., Barnes-Holmes, D., & Vahey, N. (in press). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioural Science*.

- Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record, 61*, 391–410.
- 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*, 465–496.
- 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.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697-720.
- Karpinski, A., & Steinman, R. B. (2006). The single category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16–32.
- Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76*, 131–158.
- Kenrick, D. T., & Gutierrez, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology, 38*, 131-140.

- 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 behavioural response patterns in contingency aware and unaware subjects during a picture–picture conditioning paradigm. *Neuroscience*, 158, 721–731.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2001). *International Affective Picture System (IAPS). Instruction manual and affective ratings (Tech. Rep.)*. Gainesville, FL: The University of Florida, The Center for Research in Psychophysiology.
- 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.
- LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, 37, 570-583.
- Levey, A. B., & Martin, I. (1975). Classical conditioning of human “evaluative” responses. *Behaviour Research and Therapy*, 13, 221–226.
- Levin, M. E., & Hayes, S. C. (2009). ACT, RFT, and contextual behavioural 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.
- Lionello-DeNolf, K. M., Urcuioli, P. J. (2002). Stimulus control topographies and test of symmetry in pigeons. *Journal of the Experimental Analysis of Behavior, 78*, 467–495.
- Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 97–106.
- Luciano, C., Gómez-Becerra, I., & Rodríguez-Valverde, M. (2007). The Role of Multiple-Exemplar Training and Naming in Establishing Derived Equivalence in an Infant. *Journal of Experimental Analysis of Behavior, 87*, 349-365.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology, 55*, 28-36.
- Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1*, 57-102.
- Martin, I., & Levey, A. B. (1994). The evaluative response: Primitive but necessary. *Behaviour Research and Therapy, 32*, 301–305.
- McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2007). Deictic relational complexity and the development of deception. *Psychological Record, 57(4)*, 517-531.
- McHugh, L., & Stewart, I. (2012). *The self and perspective taking: Contributions and applications from modern behavioural science*. Oakland: New Harbinger Publications.

- Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: It's not what you thought it was. *American Psychologist, 61*, 10–26.
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioural and Brain Sciences, 32*, 183–198.
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.
- Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski & K. B. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and application*. NY: Guilford Press.
- Moran, T., & Bar-Anan, Y. (in press). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion*.
- Munnely, A., Dymond, S., & Hinton, E. C. (2010). Relational reasoning with derived comparative relations: A novel model of transitive inference. *Behavioral Processes, 85*, 8–17.
- Nicholson, E., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62*, 263–278.
- Nosek, B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition, 19*, 625–664.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: a methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265–92). New York: Psychology Press.

- 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., 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.
- O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52*, 51-76.
- O'Toole, C., Barnes-Holmes, D., & Smyth, S. (2007). A derived transfer of functions and the Implicit Association Test. *Journal of the Experimental Analysis of Behavior, 88*(2), 263-283.
- 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.
- Pepper, S. C. (1942/1970). *World hypotheses: A study in evidence*. Berkeley, CA: University of California Press.
- 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.
- Power, P. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record*, 59, 621–640.
- Ratliff, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately, explicit attitude generalization takes time. *Psychological Science*, 19, 249–254.
- 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.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II* (pp. 64–99). Appleton-Century-Crofts.
- 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., & Dymond, S. (2008). A transformation of functions in accordance with the nonarbitrary relational properties of sexual stimuli. *The Psychological Record*, 58, 71-94.
- 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.
- Roddy, S., Stewart, I. & Barnes-Holmes, D. (2011). Facial reactions reveal that slim is good but fat is not bad: Implicit and explicit measures of body size bias. *European Journal of Social Psychology*, 41(6), 488-494.
- Rumbaugh, D. M., Savage-Rumbaugh, S., King, J. E., & Taglialatela, J. P. (2011). The Foundations of Primate Intelligence and Language. In D. C. Broadfield, Schick, K., Toth, N., & Yuan, M., (Eds.), *The Human Brain Evolving: Paleoneurological Studies in Honor of Ralph L. Holloway*. Stone Age Institute Press: Indiana.
- Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*, 23, 1118-1152.
- 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.
- Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence inconsistent attitudes. *Psychological Science*, 17, 954-958.
- Schwarz, N., Münkler, T. & Hippler, H. J. (1990). What determines a 'perspective'? Contrast effects as a function of the dimension tapped by preceding questions. *European Journal of Social Psychology*, 20, 357-361.

- Sidman, M. (1994). *Equivalence relations and behaviour: A research story*. Boston, MA: Authors Cooperative.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, *37*, 23–44.
- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, *52*, 270-277.
- 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.
- Smeets, P. M., Leader, G., & Barnes, D. (1997). Establishing stimulus classes with adults and children using a respondent training procedure: A follow-up study. *The Psychological Record*, *47*, 285-308.
- Smith, E.R., & De Coster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.
- 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.
- Steele, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, *56*, 519-555.

- Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *The Psychological Record, 54*, 531-552.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220–247.
- Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical Conditioning of Consumer Attitudes: Four Experiments in an Advertising Context. *Journal of Consumer Research, 14*, 334-349.
- Summerville, A., & Roese, N. J. (2008). Dare to compare: Fact-based versus simulation-based comparison in daily life. *Journal of Experimental Social Psychology, 44*, 664-671.
- Sweldens, S., Van Osselaer, S., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research, 37*, 473–489.
- Törneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical applications*. Oakland, CA: New Harbinger Publications, Inc.
- Uhlmann, E. L., Pizarro, D. A., & Bloom, P. (2008). Varieties of unconscious social cognition. *Journal for the Theory of Social Behaviour, 38*, 293-322.
- Urcuioli, P. J. (2008) Associative symmetry, “anti-symmetry”, and a theory of pigeons' equivalence-class formation. *Journal of the Experimental Analysis of Behavior; 90*, 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.

- Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, *59*, 371–388.
- Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A. & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes* *32*, 71–79.
- 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.
- Waldmann, M.R., & Hagmayer, Y. (2005). Seeing vs. doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 216–227.
- 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., & Grigoriadis, S. (2004). Why sad people like shoes better: The influence of mood on the evaluative conditioning of consumer attitudes. *Psychology & Marketing*, *21*, 755–773.
- 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.
- Wang, T., McHugh, L. & Whelan, R. (2012). A test of the discrimination account in equivalence class formation. *Learning and Motivation*, 43, 8-13.
- Watson, J. B. (1924). *Behaviorism*. New York: People's Institute Publishing Company.
- Watson, J. B. & Rayner, B. (1920). Conditioned emotional reactions. *Journal of Experimental Child Psychology*, 3, 1–14.
- 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.
- Whelan, R., Barnes-Holmes, D., & Dymond, S. (2006). The transformation of consequential functions in accordance with the relational frames of more-than and less-than. *Journal of the Experimental Analysis of Behavior*, 86(3), 317-335.
- Whitfield, M., & Jordan, C.H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology*, 45, 748–759.
- Wilson, D. S., Hayes, S. C., Biglan, T. & Embry, D. (in press). Evolving the future: Toward a science of intentional change. *Behavioural and Brain Sciences*.
- 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.
- Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, 43, 155–165.

Appendix A: Consent Form

PARTICIPANT:

I consent to participate in an experimental psychology study being run by Sean Hughes and supervised by Professor Dermot Barnes-Holmes in the Department of Psychology, National University of Ireland, Maynooth.

I understand and consent to the following:

- o The experiment will not last longer than 2 hours on any given day.
- o All data from the study will be treated confidentially.
- o The data will be stored in a locked cabinet in the Department of Psychology
- o The data will be retained for a minimum of five years.
- o An alphanumeric code will be used to protect your identity. This alphanumeric code will also be used on all measures to protect your identity.
- o Your data is available to you at your discretion
- o The data collected as part of this study will be collated and form part of Sean Hughes' doctoral thesis and the results may be included in other publications.
- o I am free to terminate my participation in the study at any time and may withdraw the data obtained from my participation, if I so wish, up to the time of publication.
- o Results from this research work will not be used deceptively or without your consent.
- o If during my participation in the study I feel the information and guidelines I have been given are neglected or disregarded in anyway, or if I am unhappy about the process I may contact the Secretary of the National University of Ireland Maynooth Ethics Committee at pgdean@nuim.ie or 01 708 6018.
- o I have been assured that my concerns will be dealt with in a sensitive manner.
- o I have received this information in an understandable way.
- o I was given at least 24 hours before agreeing to volunteer for this study.

o All my questions at this stage have been answered.

Please print and sign your name below if you are willing to abide fully by the conditions stated above.

Name:

(Please print in block capitals)

Signature:

Date: _____

EXPERIMENTER:

I, Sean Hughes, as primary experimenter, accept full responsibility for the care of all experimental participants and I confirm that all the necessary safety precautions have been taken.

Signature of experimenter: _____ Date: _____

Sean Hughes

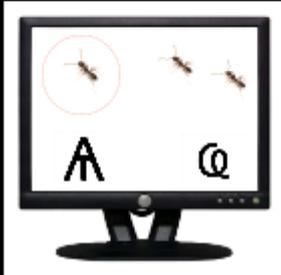
c/o Department of Psychology

NUI Maynooth

Appendix B: Contextual Cue Training Instructions

****PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY****

IN PART 1 OF THE EXPERIMENT YOU WILL SEE A SERIES OF COMPUTER SCREENS THAT EACH CONTAIN (A) TWO PICTURES AT THE TOP OF THE SCREEN AND (B) TWO SYMBOLS AT THE BOTTOM OF THE SCREEN.



YOUR TASK IN THE FIRST PART OF THE EXPERIMENT IS TO DETERMINE WHAT THESE TWO SYMBOLS MEAN



HOW DO YOU FIGURE OUT WHAT THESE TWO SYMBOLS MEAN?

****BY LOOKING AT THE RELATIONSHIP BETWEEN THE PICTURES AT THE TOP OF THE SCREEN****



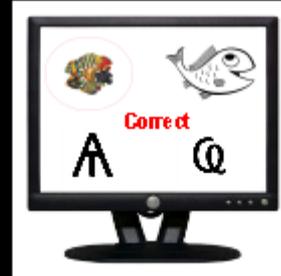
FOR EXAMPLE, IF YOU CHOOSE THE SYMBOL ON THE RIGHT USING THE '@' KEY THEN YOU ARE SAYING THAT THE SYMBOL ON THE RIGHT DESCRIBES THE RELATIONSHIP BETWEEN THE PICTURES.



IF YOU CHOOSE THE SYMBOL ON THE LEFT USING THE 'A' KEY THEN YOU ARE SAYING THAT THE SYMBOL ON THE LEFT DESCRIBES THE RELATIONSHIP BETWEEN THE PICTURES.



THE COMPUTER WILL HELP YOU BY TELLING YOU WHEN YOU ARE RIGHT AND WHEN YOU ARE WRONG. USE THIS FEEDBACK TO HELP YOU LEARN.

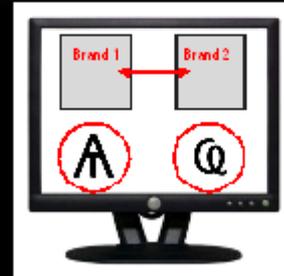


Appendix C: Relational Training Instructions

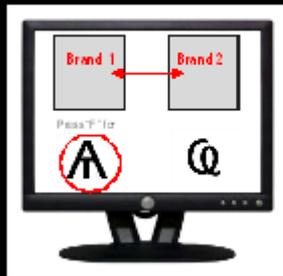
IN PART 2 YOU WILL SEE A SERIES OF COMPUTER SCREENS THAT EACH CONTAIN (A) THE NAMES OF TWO BRAND PRODUCTS AND (B) THE TWO SYMBOLS YOU JUST LEARNED ABOUT.



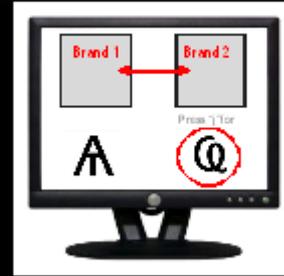
YOUR JOB IS TO FIGURE OUT WHAT THE RELATIONSHIP IS BETWEEN THE TWO BRAND PRODUCTS AT THE TOP OF THE SCREEN USING THE SYMBOLS YOU HAVE JUST LEARNED ABOUT.



IF YOU CHOOSE THE SYMBOL ON THE LEFT, YOU ARE SAYING THAT THE RELATIONSHIP BETWEEN THE TWO BRAND PRODUCTS IS WHAT THAT SYMBOL MEANS.

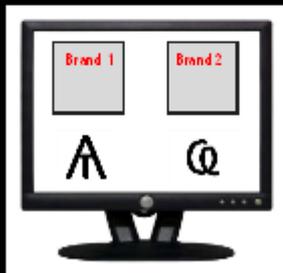


IF YOU CHOOSE THE SYMBOL ON THE RIGHT, YOU ARE SAYING THAT THE RELATIONSHIP BETWEEN THE TWO BRAND PRODUCTS IS WHAT THAT SYMBOL MEANS.



REMEMBER

THE AIM IN PART 2 TO FIGURE OUT THE RELATIONSHIP BETWEEN THE TWO BRAND PRODUCTS USING WHAT YOU HAVE LEARNED THE TWO SYMBOLS TO MEAN.



Appendix D: Stimulus Rating Task

Based on what you have learned throughout this experiment, please rate how positive or negative you feel towards the brand product below using the slider provided.

-4 = Negative Feelings, 0 = Neutral, 4 = Positive Feelings

Pardal

Negative Neutral Positive

-4 -3 -2 -1 0 1 2 3 4

Press button to continue

Appendix E: Contextual Cue Rating Task

USING THE BOX BELOW, PLEASE INDICATE
WHAT YOU THINK EACH OF THE SYMBOLS MEAN

	
<input data-bbox="513 716 777 792" type="text"/>	<input data-bbox="858 716 1121 792" type="text"/>

PRESS TO CONTINUE

Appendix F: Demand Compliance Task

YOU JUST RATED HOW MUCH YOU LIKED OR DISLIKED THE BRAND NAMES.
DID YOU BASE HOW MUCH YOU LIKED OR DISLIKED THE BRAND NAMES
ON WHAT YOU THOUGHT THE RESEARCHER WANTED YOU TO DO OR
BASED ON WHAT YOU LEARNED ABOUT THEM?

- I based my response on what I thought the researcher wanted me to do
- I based my response on what I learned about the brand names
- I don't know why I evaluated the brand names as I did

press to continue