Journal of Contextual Behavioral Science 1 (2012) 17–38

Contents lists available at SciVerse ScienceDirect

Journal of Contextual Behavioral Science

journal homepage: www.elsevier.com/locate/jcbs

Review Paper

Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research

Sean Hughes *, Dermot Barnes-Holmes, Nigel Vahey

Department of Psychology, National University of Ireland Maynooth, Ireland

ARTICLE INFO

Article history:
Received 29 June 2012
Received in revised form
20 September 2012
Accepted 25 September 2012

Keywords:
Implicit cognition
Functional contextualism
IRAP
REC model

ABSTRACT

Throughout much of the past two decades, contextual behavioral science has been applied to a diverse spectrum of psychological phenomena. This intellectual voyage into uncharted waters has brought with it exciting new developments at the methodological and theoretical levels as well as increased contact with other philosophical frameworks such as mechanism. This expansion into new territories requires that the researcher carefully walk a tight-rope between different intellectual traditions—an activity that is subject to several challenges and dangers. In the following paper we provide a detailed map on how to navigate such pitfalls in the study of implicit cognition. We open with a comprehensive overview of the core assumptions and analytic strategies upon which the cognitive (mechanistic) and functional (contextual) traditions have been built. As we shall see, both traditions have sought to understand, predict, and in some cases influence, behavior using radically different conceptual, theoretical and methodological tools. The Relational Elaboration and Coherence (REC) model as well as the Implicit Relational Assessment Procedure (IRAP) are offered as examples of how researchers can explore the domain of implicit cognition from a purely functional perspective. Finally, we examine the possibility that although the cognitive and functional frameworks operate at two independent levels of analysis each may be mutually informed by the work of the other, to the benefit of both.

© 2012 Association for Contextual Behavioral Science. Published by Elsevier Inc. All rights reserved.

Contents

1. Introduction . . . 18
2. The mechanistic approach to psychological science . . . 18
3. The mechanistic approach to implicit cognition . . . 19
   3.1. Operating processes and conditions . . . 19
   3.2. Learning history and current context . . . 20
   3.3. Procedural properties (construct relevant) . . . 20
   3.4. Procedural properties (construct irrelevant) . . . 21
   3.5. Conclusion . . . 21
4. Challenges facing the mechanistic approach . . . 22
5. A functional approach to psychological science . . . 22
   5.1. Relational frame theory . . . 23
6. A functional approach to the study of implicit cognition . . . 24
7. The relational elaboration and coherence (REC) model . . . 24
   7.1. Low complexity low derivation responding . . . 26
      7.1.1. The IRAP . . . 26
   7.2. Higher complexity low derivation . . . 27
      7.2.1. The problem with EERRs . . . 28
   7.3. Low complexity higher derivation . . . 29
   7.4. High complexity high derivation . . . 29
   7.5. Conclusion . . . 30
8. BIRRs and EERRs: the role of relational coherence, predictive validity and awareness . . . 30
   8.1. Relational coherence . . . 30

* Corresponding author.
E-mail addresses: sean.hughes@nuim.ie (S. Hughes), dermot.barnes-holmes@nuim.ie (D. Barnes-Holmes).

2212-1447/$ - see front matter © 2012 Association for Contextual Behavioral Science. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcbs.2012.09.003

8.2. Prediction of behavior . . . 31
8.3. The "awareness" issue . . . 31
9. Implications of the REC model . . . 32
9.1. General conceptual issues . . . 32
9.2. Levels of derivation rather than verbal versus non-verbal . . . 32
9.3. Limitations and future directions. . . . 33
10. Epilogue . . . 34
Acknowledgments . . . 34
References . . . 34

1. Introduction

Psychological science has long sought to unravel the enigma of how the inner and private world of the individual interacts with their overt behavior (for reviews see Banaji & Heiphetz, 2010; Nosek, Hawkins, & Frazier, 2011; Payne & Gawronski, 2010). More often than not, this work has been inspired by finding that people behave in two different and potentially conflicting ways. On the one hand, and consistent with our intuitive beliefs about behavior, we can respond to stimuli in our environment in a non-automatic fashion. These "explicit" responses are argued to be controlled, "intentional, made with awareness and require cognitive resources" (Nosek, 2007, p. 65) and are typically registered using direct measurement procedures such as questionnaires, interviews, and focus groups. On the other hand, our history of interacting with the social environment can also result in the formation of automatic or "implicit" responses that are characterized by differing degrees of awareness, intention and control (see Gawronski & Payne, 2010) and are often assessed using indirect procedures 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 Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006; for an overview of these tasks see Nosek et al., 2011; Gawronski & De Houwer, in press).[1]

An extensive literature now indicates that automatic and controlled responses correspond with one another when phobic stimuli (Teachman, 2007), consumer preferences (Maison, Greenwald, & Bruin, 2004) and political orientation (Choma & Hafer, 2009) are subject to inquiry. However, under specific conditions – and with respect to certain stimuli – automatic and controlled responding may also conflict. For instance, people often show automatic negative responses to members of other racial, ethnic or religious groups despite their self-reported egalitarian sentiments (McConnell & Leibold, 2001; Payne, Burkley, & Stokes, 2008). Automatic responses may not only correspond or diverge from controlled behaviors but – perhaps more importantly – differentially, additively or interactively predict future actions over and above their non-automatic counterparts (see Perugini, Richetin, & Zogmaister, 2010). For example, the degree to which people seeking psychiatric treatment automatically relate the self with death/suicide predicts their likelihood of attempting suicide in the following 6 months (Nock et al., 2010), while automatic evaluative responding to one's partner predicts risk of future relationship breakup (Lee, Rogge, & Reis, 2010). Likewise, the quality and quantity of social interactions with members of other racial or ethnic groups (McConnell & Leibold, 2001), voting behavior (Galdi, Arcuri, & Gawronski, 2008) and likelihood of engaging in safe sex behaviors (Czopp, Monteith, Zimmerman, & Lynam, 2004) have all been predicted on the basis of automatic responding.

As in any area of (psychological) science, researchers interested in the study of implicit cognition have adopted a set of philosophical assumptions about the research domain, appropriate units of analysis and relevant truth criteria. When taken together, these pre-analytic assumptions provide the philosophical scaffold upon which individual theories have been built, methodologies crafted and empirical findings interpreted (see Hughes, Barnes-Holmes, & De Houwer, 2011). Although a number of philosophical frameworks or "worldviews" have been proposed to guide scientific activity (Pepper, 1942; Hayes, Hayes, & Reese, 1988), research in this area has been dominated by psychologists subscribing to a cognitive (mechanistic) – and to a lesser degree – functional (contextual) position (referred to hereafter as the mechanistic and functional approaches).

In the following paper we open with a detailed overview of the core assumptions and analytic strategies upon which the mechanistic and functional traditions have been built. Clarifying the goals of science, nature of truth and the basic units that different researchers adopt in the study of implicit cognition will provide the reader with an essential background against which to evaluate the rationale for past and future developments within this research area. Although the mechanistic approach has long guided this research enterprise we offer the Relational Elaboration and Coherence (REC) model as well as the Implicit Relational Assessment Procedure (IRAP) as examples of how researchers can explore this intellectual domain from a purely functional perspective. Importantly, this expansion into new territories requires that the researcher carefully walk the tight-rope between the mechanistic and functional traditions – an activity that is subject to several challenges and dangers. We outline how researchers can avoid these problems by maintaining a firm separation between the mechanistic and functional levels of analysis. Finally, and in line with recent work by De Houwer (2011), we close by examining the possibility that although the mechanistic and functional frameworks operate at two independent levels of analysis each may be mutually informed by the work of the other, to the benefit of both.

2. The mechanistic approach to psychological science

Arguably, the majority of empirical work within contemporary psychology – and implicit cognition in particular – has been conducted by researchers operating within a mechanistic worldview. Broadly speaking, mechanists conceptualize (psychological) events as being similar to a machine, composed of discrete parts that interact and are subject to specific operating conditions

[1] Following De Houwer (2008), we define a procedure as either *direct* or *indirect* on the basis of its procedural properties and the outcome or effect of a procedure as either *implicit* or *explicit* based on the properties of the processes by which the attribute influences behavior. Although cognitive researchers often use terms such as implicit and explicit to distinguish conscious from unconscious (Greenwald & Banaji, 1995) or automatic from controlled processes (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009), functional researchers employ these same terms simply as descriptive labels for different classes of behavior rather than mediating mental mechanisms. That said, several researchers have recently attempted to provide a functional definition of automaticity (De Houwer, & Moors, in press).

(Bechtel, 2008). The goal of psychological science from this perspective is twofold. First, the basic mental processes that mediate between input (environment) and output (behavior) 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 behavior. 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 behavioral observations that it aims to predict. Put simply, mechanistic researchers are primarily focused on the prediction of behavior through the use of theoretical models that bridge past events and current responses. Such models are usually high in precision and low in scope. For example, a mental model may specify in elaborate detail the operating processes and conditions responsible for implicit cognition yet remain silent to other cognitive phenomenon such as language, intelligence or perspective taking (see Gawronski & Creighton, in press).[2]

At the same time, mechanistic researchers must also identify the operating conditions that are both necessary and sufficient for mental processes to successfully function. Stated more precisely, mental processes are argued to operate under a restricted set of conditions and these are separate from, but co-vary with, the environmental context under which behavior is observed (Moors & De Houwer, 2006; Moors, Spruytm, & De Houwer, 2010). From 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 operating conditions necessary for those processes to occur. For instance, many researchers 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 behavior (e.g., efficient, unaware, uncontrollable, unintentional). Likewise, procedural properties that typically characterize direct procedures such as an absence of speed or accuracy criteria, high task complexity or the requirement to verbally report activity, are argued to correspond to the operating conditions necessary for non-automatic mental activity and thus non-automatic behavior (e.g., non-efficient, aware, controllable and intentional; see De Houwer et al., 2009).[3] In other words, "the correspondence between the conditions established by a certain kind of measurement procedure (indirect vs. direct) and the operating conditions of a certain kind of mental process (automatic vs. controlled) is a fundamental assumption" for mechanistic researchers (Peters, 2011, p. 6).

To summarize, when a mechanistic framework is adopted, behavioral causation is the product of mental mediation. Understanding the world therefore requires that the basic units that fill the temporal gaps between one event and another be identified. This assumption serves to orientate empirical attention away from environment–behavior interactions and towards proximal causes that are assumed to generate the behavior obtained on a given procedure. The result is a largely hypothetico-deductive and theory-driven research agenda. As we shall see, this mechanistic world-view has shaped the theoretical and methodological landscape of implicit cognition research for several decades now.

3. The mechanistic approach to implicit cognition

Although there are non-trivial differences across mental models of implicit cognition, the assumption that an individual's learning history and current context indirectly influence the behavior obtained on a measurement procedure – through the operation of a set of mental processes and conditions – is central to many, if not most mechanistic accounts (see Fig. 1). In the following sections we consider the implications and challenges associated with explaining different patterns of behavior in-line with this pre-analytic assumption.

3.1. Operating processes and conditions

In light of the fact that automatic responding can correspond or conflict with non-automatic behavior, as well as predict important real-world outcomes, increased attention has been paid to how this particular class of responses is established, manipulated, and changed. For cognitive (mechanistic) researchers this has involved constructing and testing theories to explain how mental processes and conditions interact with one another to guide implicit and explicit cognition (e.g., Albarracin, Glasman, & Wallace, 2004; Fazio, 2007; Gawronski & Bodenhausen, 2011; Greenwald et al., 2002; Lieberman, Gaunt, Gilbert, & Trope, 2002; Petty, Briñol, & DeMarre, 2007; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004). These various accounts differ in the importance they allocate to specific mental processes/representations and the conditions under which they function.

What is important to note here is that irrespective of the particular construct, the vast majority of models conceptualizes automatic and controlled responding as the product of proximal mental events. Whether framed as a single process account involving associations (Fazio, 2007; Petty et al., 2007), a dual process model involving reflective-impulsive systems (Cunningham, Zelazo, Packer, & Van Bavel, 2007; Strack & Deutsch, 2004), a combination of associations and propositions (Gawronski & Bodenhausen, 2011; Rydell & McConnell, 2006), or multiple interactive memory systems (Amodio & Ratner, 2011; Smith & DeCoster, 2000), the idea that contiguous mental mechanisms mediate between environment and behavior is foundational. At the same time, the above models also share the

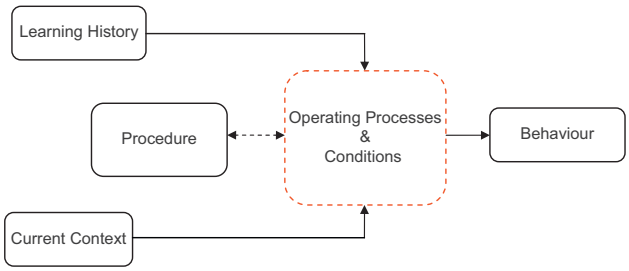

Fig. 1. Behavior causation according to the cognitive (mechanistic) framework. The effects of learning history and wider context impact on behavior indirectly through a set of mediating (mental) operating processes and conditions. Construct relevant and irrelevant properties of the procedure determine what operating conditions and processes are assumed to be causally related to behavior and whether that behavior is deemed automatic or controlled. Note that while the procedure is a feature of the current context, we highlight it here given its explanatory importance within the mechanistic literature.

[2] It should be noted that mechanistic researchers distinguish between various types of mental mediators, such that some fundamental mental concept (e.g., representation) is acted upon by a mental process (e.g. spread of activation, truth validation) and these mental processes in turn operate under different mental conditions (e.g., conscious, efficient, intentional and controllable; see Gawronski & Bodenhausen, 2011).

[3] Clarifying the relationship between operating processes and conditions in the study of implicit cognition is made all the more difficult given that automaticity is often defined in terms of a number of subcomponents that different mental processes may or may not demonstrate in certain contexts and different procedures may or may not target depending on their particular properties (De Houwer et al., 2009).

assumption that mental processes often occur under mental operating conditions referred to the "four horsemen of automaticity": awareness, intentionality, efficiency, and control (Bargh, 1994; De Houwer & Moors, 2010).[4] For instance, some authors argue that the degree to which automatically activated associations guide "downstream" overt behavior depends on two operating conditions: the cognitive resources available to the person and their motivation and opportunity to deploy them (Olson & Fazio, 2009; Petty et al., 2007). For others, behavior is rooted in two conceptually distinct kinds of processes (i.e., associative linking vs. propositional reasoning) that operate under different subcomponents of automaticity (Gawronski & Bodenhausen, 2011).

In short, mechanistic theories are primarily concerned with how mental constructs are formed, activated and changed as well as influence subsequent behavior. With respect to implicit cognition, this involves identifying the operating processes that suggest *what* a given construct is doing (e.g., associative activation, propositional reasoning) and the operating conditions that indicate *when* that process takes place (e.g., when cognitive resources and time are restricted or when the person intends to manipulate the process or its outcome).

3.2. Learning history and current context

Although an individual's learning history and the current context influence the behavioral outcomes obtained from direct and indirect procedures, they do so distally through their influence on the aforementioned operating processes and conditions. On the one hand, a history of learning with respect to particular stimuli may extend backwards in time to early socialization experiences (Rudman, Phelan, & Heppen, 2007; Sinclair, Dunn, & Lowery, 2005) or involve more recent social interactions (Shook & Fazio, 2008), descriptive information (Gawronski, Rydell, Vervliet, & De Houwer, 2010) or contingencies established within the laboratory (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; for a review see Gawronski & Sritharan, 2010). In each case, however, the learning processes and boundary conditions that influence the behavioral change are of interest only in so far as they specify how and when mental processes/representations function. For example, interactions between the person and environment in the form of respondent learning (Olson & Fazio, 2001), operant learning (Peters & Gawronski, 2011), stimulus generalization (Ranganath & Nosek, 2008) or verbal instruction (Gregg, Seibt, & Banaji, 2006) are conceptualized as influencing the "strength" of mental associations between stimuli in memory and it is these associations that govern the subsequent response (see also Rydell & Gawronski, 2009). Specifically, the strength of stored associations is often assumed to dictate the degree to which previously encoded information is automatically reactivated when a stimulus is encountered and an automatic response emitted.[5]

On the other hand, the environment – whether in the form of the current or historical context – is treated as an object such that the person's actions occur in it but the causes of that behavior are separate from it. More often than not, cues in the environment are argued to differentially influence the activation of mental associations in memory and the behavior resulting from such activity. Consider, for example, the well-replicated finding that automatic evaluative responding towards African Americans differs according to the context in which they are presented (e.g., church-interior versus street corner; Wittenbrink et al., 2001). Several researchers propose that the mental associations established by the person's learning history are highly stable and resistant to changes across time and context (Olson & Fazio, 2009; Petty et al., 2007). Thus responding more positively to African Americans when depicted in a church – and more negatively in an urban setting – is understood not in terms of variations in context but instead as differences in how stimuli are categorized or whether responding reflects the endorsement of extra-personal or personal associations. In such cases, features of the current context are thought to contaminate the assessment of a person's "true" and enduring evaluative disposition as well as limit the ability to predict future behavior. In contrast, others argue that automatic evaluative responding is acutely sensitive to changes in context, with different patterns of associations momentarily activated as a function of the cues present in the environment (Ferguson & Bargh, 2004; Gawronski & Bodenhausen, 2011; Schwarz, 2007). According to these accounts, encountering a stimulus within a given context (e.g. black person in a church) only serves to activate a limited number of all the possible associations available in memory related to that stimulus. In this way the same stimulus will elicit different evaluative responses depending on the context in which it is encountered and the pre-existing associations in memory.

In short, variation in automatic responding across different situations is often explained in one of two ways. The first views the associations established through the person's interaction with their environment as "context-insensitive"; highly durable and robust to change. The second view is a "context-sensitive" one in which a subset of mental associations are activated when the individual's learning history interacts with the current context.

3.3. Procedural properties (construct relevant)

Although direct and indirect procedures represent features of the current context, they have attracted considerably more attention within the mechanistic literature than many other environmental variables relevant to implicit cognition. Broadly speaking, a procedure constitutes a systemic structuring of an environmental situation aimed at generating a behavioral outcome. "It specifies, among other things, the way in which stimuli should be presented, responses registered, and the outcome derived from those responses" (De Houwer, 2007, p. 231). Given that mechanistic researchers are focused on discovering the "true" mental construct(s) that mediate between environment and behavior, the effect obtained with a procedure is typically treated as the causal product of variations in the underlying mental attribute or process of interest. To illustrate this more clearly, consider two different indirect procedures commonly

[4] Although operating conditions such as intention, control, efficiency and awareness refer to properties of the mental construct under investigation (and thus why we placed them inside the "black-box" in Fig. 1) it is possible to define these various features of automaticity purely in terms of observable elements in the environment (De Houwer & Moors, in press). When treated in this way, operating conditions no longer reside within the mental "black-box" but represent features of the past and present context. Note, however, that this "functional decompositional" approach to automaticity does not reflect the established orthodoxy within the mechanistic literature (e.g., Gawronski & Bodenhausen, 2011).

[5] Mechanistic researchers often treat explicit evaluations as the product of "propositional learning processes" and implicit evaluations as the product of "associative learning processes" (e.g., Peters & Gawronski, 2011; McConnell, Rydell, Strain & Mackie, 2008). That said, it is not entirely clear to what degree associative and propositional learning processes involve or at least correspond to

(*footnote continued*)
the operation of respondent, operant or relational learning as functionally understood.

used to target automatic responding towards body-size; the IAT and evaluative priming.[6]

On the IAT a target (e.g., Average Weight) and attribute stimulus (e.g., Good) are presented in close spatial proximity on one side of a computer screen and a second target (Over Weight) and attribute (Bad) on the other. Participants are presented with a successive stream of positive and negative words as well as pictures of average weight and obese individuals in the middle of the screen and required to categorize those stimuli with their respective category labels. More rapid and accurate responding when stimuli are mapped to the same response option (*Over Weight–Bad; Average Weight–Good*) versus different response options (*Over Weight–Good; Average Weight–Bad*) allows for an inference to be made about how those stimuli are automatically associated in memory (e.g., Teachman & Brownell, 2001). Similarly, participants completing an evaluative priming task encounter a sequence of trials in which a prime (picture of an average weight person) and target stimulus (Good) are presented in close temporal proximity. During the task only the target has to be processed along an evaluative dimension while the prime is treated as an irrelevant distractor. The speed and/or accuracy with which participants respond to the prime-target relationship across different conditions is treated as a proxy for the automatic activation or processing of the prime stimulus (Cserjesi et al., 2010).

In other words, although the IAT and priming tasks differ in several notable ways, the observable outcome resulting from a core property of both procedures (i.e., the temporal or spatially contiguous presentations of stimuli) is often treated as the causal product of a particular mental process (i.e., activation of mental *associations* in memory). This mechanistic assumption underpins not only the IAT and evaluative priming but nearly every indirect procedure constructed to date, including the Stimulus Response Compatibility Task (SRCT; Mogg, Bradley, Field, & De Houwer, 2003), Implicit Association Procedure (Schnabel, Banse, & Asendorpf, 2006), Go/No-go Association Task (Nosek & Banaji, 2001) and Affective Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005). Irrespective of their idiosyncratic features (e.g., physical pairing of stimuli, approach versus avoidance responding) the behavioral effects generated by these measurement tools are taken to correspond to an underlying mental mechanism (see Gawronski & De Houwer, in press).

Despite this assumption, a growing body of research indicates that the behavioral outcomes obtained from indirect procedures do not provide a one-to-one correspondence with associative/heuristic/impulsive processes nor do performances on direct procedures map directly onto propositional/systematic/reflexive processes. For instance, "associative" measures such as the IAT and priming are sensitive to propositional learning involving stimuli that have never been directly paired (e.g., De Houwer, 2006b; Gast & De Houwer, 2012). Similarly, propositional and associative learning influence the behavioral outcomes obtained on both direct and indirect procedures (Whitfield & Jordan, 2009). Taken together, these findings suggest that no procedure provides a "process-pure" picture of the mental construct of interest or the conditions of processing. Rather, for mechanistically orientated researchers, the behavior obtained on such tasks reflects the joint contribution of multiple mental processes that may be characterized by different features of automatic or controlled processing.

3.4. Procedural properties (construct irrelevant)

At the same time, the behavioral outcomes obtained on indirect procedures may also be governed by properties of the measure that are independent of the mental construct under investigation. For tasks involving a block structure and alternating response options (such as the IAT) stimulus-response compatibility (De Houwer, 2001), task switching (Klauer & Mierke, 2005), figure-ground asymmetry (Rothermund & Wentura, 2004), recoding (Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009), and block order (Nosek, Greenwald, & Banaji, 2007) may influence not only the absolute size but potentially the rank order of obtained effects. Likewise, intentionally rating the primes instead of the targets (Bar-Anan & Nosek, 2012), attention to category membership (Gawronski, Cunningham, LeBel, & Deutsch, 2010), length of stimulus onset asynchrony and backward checking strategies represent sources of construct irrelevant variance on priming tasks (see Wentura & Degner, 2010 for a detailed discussion).[7] More generally, the absence of a non-arbitrary zero point (Blanton & Jaccard, 2006), ability to exert strategic control over the behavioral effect (Fiedler & Bluemke, 2005) as well as individual differences in cognitive ability have been shown to influence the findings obtained from a number of different indirect procedures (see Teige-Mocigemba, Klauer, & Sherman, 2010). Consequently, task-specific properties that characterize one indirect procedure but are absent from another may explain why different tasks sometimes show opposite or alternative outcomes for theoretically comparable manipulations (Deutsch & Gawronski, 2009). Overall, these findings provide further evidence that the effects obtained from direct and indirect procedures do not provide a one-to-one correspondence with the mental constructs that they are designed to assess. Rather the unique constellation of properties that characterize a procedure will dictate what combination of operating processes and conditions will be brought to bear in the measurement context, as well as the potential sources of systematic error variance that mediate between those processes and the observed outcome.

3.5. Conclusion

In short, the mechanistic approach has guided the study of implicit cognition for well over 20 years now, shaping the questions asked, methodologies devised and theoretical interpretation offered. The success of this enterprise is evident in the widespread adoption of these methods and theories across a diverse spectrum of research domains, including health psychology (Wiers et al., 2010), consumer psychology (Perkins & Forehand, 2010), forensic psychology (Snowden, Craig, & Gray, 2011) and clinical psychology (Teachman, Cody, & Clerkin, 2010). However, as Hayes (2004) notes "when a discipline is markedly successful, it tends to continue in the same direction for a time without a serious examination of its assumptions because adherents have interesting work to do and rewards for doing that work. Eventually, however, these assumptions begin to be examined". Consistent with this notion, a number of authors have

[6] For cognitive mechanistic researchers an outcome generated by a measurement procedure is often defined as "implicit" when it is assumed to be causally produced by a psychological attribute or construct that operates under one or more of the conditions of automaticity. Consequently, different procedures may produce effects that are "implicit" in different ways depending on (i) how their task-specific properties co-vary with the different operating conditions of automaticity and (ii) according to the automaticity features that characterize the metal process brought to bear within that measurement context (see De Houwer & Moors, 2010).

[7] The discovery of method related variance on IAT effects has led to the emergence of a second generation of IAT alternatives that aim to circumvent the block structure of the task and its relativistic nature. Examples include the Single Block IAT (Teige-Mocigemba, Klauer, & Rothermund, 2008), Recoding Free IAT (Rothermund et al., 2009) and the Sorting Paired Features Task (Bar-Anan, Nosek, & Vianello, 2009). Equally, a number of strategies have also been devised to reduce method variance on sematic and evaluative priming tasks (see Wentura & Degner, 2010).

recently argued that, despite its centrality to contemporary psychological research, mechanism is subject to a number of challenges at the measurement and conceptual levels (Borsboom, Mellenbergh, & van Heerden, 2004; Chisea, 1994; De Houwer, 2011).

4. Challenges facing the mechanistic approach

As we have seen, applying the mechanistic framework to behavioral causation involves postulating hypothetical constructs that mediate between environment and behavior. This approach is immediately complicated by the fact that these constructs are independent of the physical system that generates them—yet they are assumed to interact with it. Consequently, their availability for direct manipulation is severely restricted. Mechanistic researchers have attempted to square this circle by treating the behavioral outcome obtained on a measurement procedure (e.g., IAT effect) as equivalent to the presence of the assumed construct of interest (e.g., mental associations in memory). However, this analytic strategy is problematic for one important reason: the use of behavior as a proxy for mental constructs serves to conflate the event that must be explained (i.e., behavior) with the concept used to explain that event (i.e., the mental construct), and in doing so violates a core feature of scientifically acceptable explanations (see De Houwer, 2011 for a detailed discussion). Indeed, the use of behavioral proxies appears to unleash a Pandora's Box of complications for mechanistic research and theory in general.

One the one hand, the correspondence between behavior and its underlying construct requires an inferential leap that is open to potential error. For instance, any behavioral effect that can be produced by a second mental construct in the absence of the originally proposed construct results in that effect becoming an insufficient basis on which to formulate firm conclusions about the originally proposed construct. This is an important issue given that variations in a mental construct do not necessarily correspond to variations in a behavioral outcome—the latter may be produced by other construct relevant (e.g. other mental processes) or construct irrelevant factors (e.g., procedural properties). To illustrate, consider evidence supporting the idea that IAT and priming effects can be generated in the absence of any associative pairings between stimuli (Gast & De Houwer, 2012). Such findings introduce concerns about whether automatic responding is the product of the originally proposed construct (associations) or an additional construct (propositions). At the same time, and as noted above, performance on any measurement procedure cannot provide a one-to-one correspondence with the construct it is designed to capture. Rather, the specific constellation of construct relevant and irrelevant features of the task will mediate the link between the assumed construct and the obtained behavior. "Needless to say, if effects that are driven by alternative sources of variance are misattributed to the psychological construct a procedure was designed to assess, theorizing about that construct can be seriously distorted" (Gawronski, LeBel, Peters, & Banse, 2009, p. 370). In other words, theoretical claims about underlying mental constructs and their properties are weakened in any situation that introduces uncertainty about the bi-conditional relation between behavior and the construct of interest (see Borsboom et al., 2004).

On the other hand, treating a specific mental construct as a necessary precondition for a behavioral effect may also result in empirical myopia and impede scientific progress within a given research area. For instance, the *a priori* assumption that implicit cognition is fundamentally associative in nature has become so woven into the fabric of this research area that it is often treated as an immutable truism that no longer requires investigation. Consequently, the opportunity for novel theoretical and methodological expansion that deviates from this position is often hampered. Indeed, the constraining effect of such an approach is evident in the fact that only now, after more than 20 years, the role of associations in implicit cognition has been subject to challenge (see Hughes et al., 2011 for a detailed treatment).

Finally, rather than promoting a convergence towards a coherent, overarching and "Unified Theory of Cognition" (see Garcia-Marques & Ferreira, 2011) the "behavior-as-proxy" approach has facilitated the emergence of an ever growing number of mentalistic (and often dichotomous) theories about narrow empirical phenomena. This approach seems to offer little means of selecting from between these alternative constructs. In the context of implicit cognition, for example, models are typically evaluated based on their ability to account for the domain of behavioral outcomes obtained as well as their ability to predict future behavior. Nevertheless, in situations where a number of alternative theories – each postulating different mental mechanisms – are broadly similar in their predictive and heuristic value, there appears to be no way of selecting one theory from another. In such cases, debates between competing positions cannot be resolved on empirical grounds, because any finding predicted by one account can be reinterpreted by the other (see Schwarz & Bohner, 2001). While parsimony may be called upon as an additional selection criterion it is not without its own controversies (Mitchell, De Houwer, & Lovibond, 2009).

Although researchers may be fully aware of the drawbacks in treating behavior as a proxy for mental constructs, they might continue to do so if they believe that it is the only means to study complex behaviors such as language and cognition. Critically, however, mechanism is only one among a number of philosophical frameworks that scientists can draw upon to guide their research activity. In what follows we outline a pragmatic and non-dualistic alternative – functional contextualism – and illustrate how it equips (implicit cognition) researchers with a means to circumvent many of the challenges outlined above.

5. A functional approach to psychological science

While various forms of behaviorism have emerged over the last hundred years, the most empirically and theoretically productive contemporary branch is that of contextual behavioral science (CBS; Hayes, Barnes-Holmes, & Wilson, 2012). This intellectual position is founded on an intensely pragmatic and contextual philosophy of science known as functional contextualism (Biglan & Hayes, 1996; Hayes & Brownstein, 1986) which, similar to any world-view, is characterized by a number of pre-analytic assumptions. First, this approach adopts a single scientific goal; to understand, predict and influence behavior with scope (explain a comprehensive range of behaviors 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). What differentiates this approach from mechanism is that it adopts an exclusively functional epistemology. Instead of locating behavioral causation in the mind, scientific analysis is focused on the functional relations between the (past and present) environment and behavior 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 behavioral change is omitted. There is no mechanistic requirement for spatially and temporally contiguous events to mediate the relationship between environment and behavior. Rather, behavior is defined as an on-going "act-in-context" that always occurs within and in response to a current and historical context (Pepper, 1942; Hayes, 1993). 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 behavior of interest.

A second assumption of functional contextualism as applied to behavior is that the environmental context can not only extend backwards or forwards in space and time but also may refer to the "internal environment" that occurs privately inside the person's skin or the "external environment" that takes place publicly outside the skin. Although early methodological behaviorists such as Watson (1924) focused exclusively on public behavior and excluded private events from legitimate analysis, contemporary functional researchers simply arrange behavior along a continuum from public (e.g. walking, painting) to private (e.g. thinking, feeling, and remembering) (Hayes & Brownstein, 1986; Skinner, 1945). This strategy is adopted given that prediction and influence constitute the primary goal of analysis. To illustrate this point more clearly, consider the example of a mental link between two associations in memory. Although such a mental or non-physical event may be seen as the cause of a particular behavior (performance on an IAT) it is not open to direct manipulation. In contrast, explaining a behavioral outcome in terms of environmental events such as antecedents and consequences provides the researcher with an opportunity to directly influence the causal factors that govern behavior and its change. Consequently, when scientific analysis allows for the use of mental constructs that cannot be directly manipulated, prediction appears to emerge as the primary focus rather than prediction-and-influence (Hayes & Brownstein, 1986). While prediction alone is satisfactory within the mechanistic framework, it is unsatisfactory when measured against the goals of functional contextualism. Thus by recasting emotion, (implicit) cognition and other mental constructs as a tendency to publicly or privately behave in a certain way within a certain context, scientific activity shifts from the search for mental processes and conditions towards identifying the environmental variables that govern different patterns of behavior. Put differently, functional researchers treat private events as subject to the same principles and contingencies as their public counterparts, and thus (in principle) open them up to both prediction and influence.[8]

Finally, functional researchers seek to abstract out basic findings into general and overarching "behavioral principles" that demonstrate high precision, scope and depth (e.g., the principles of reinforcement, punishment, and stimulus generalization). Such principles are inductive in nature, built from the "bottom up" and "apply across a broad array of topographically distinct behaviors of varying complexity while maintaining coherence and parsimony" (Levin & Hayes, 2009, p. 6). When researchers seek to explain a range of behavioral interactions using a set of inter-related behavioral principles, functional theories emerge. Such theories do not postulate mediating mental constructs or their interaction with other constructs when explaining behavior and its change. Instead they describe and seek to explain observed regularities in the functional relations between environmental events and behavior. One functional theory that attempts to account for a wide variety of private and public behaviors using only a handful of inter-related principles is Relational Frame Theory (Hayes, Barnes-Holmes, & Roche, 2001). In the sections that follow the basic features of this functional account will first be outlined. Thereafter, we examine how it has recently been applied to the study of implicit cognition, and in particular, how it has facilitated theoretical, methodological and empirical progress within this research area.

5.1. Relational frame theory

At its core, RFT argues that humans and non-humans alike can learn to relate stimuli (i.e., respond to one stimulus in terms of another) through a number of different learning processes, such as 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 that is previously encountered in its learning history. For instance, animals can be trained to form relations between stimuli based on their (non-arbitrary) physical properties (Harmon, Strong, & Pasnak, 1982; Reese, 1968) or relate arbitrary stimuli such as abstract shapes and symbols based on their shared functions (Vaughan, 1988). 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. In each of the above cases, relating is based on a direct history of training and often involves a specific temporal order, or formal relationship, between the to-be-related stimuli (see Hayes et al., 2001 for a detailed discussion).

In contrast, RFT argues that humans are not limited to direct contingency learning but can also relate stimuli to one another in the absence of any direct reinforcement for doing so. Put another way, while non-human animals appear to be largely restricted to relating stimuli through the use of direct contingencies and based on their formal properties, humans demonstrate the additional ability to derive novel, untrained and bi-directional relations between stimuli. RFT proposes that this behavior (termed arbitrarily applicable relational responding) is an overarching type of operant response that is learned early on in our development through interactions with the verbal community (Hayes et al., 2001; Rehfeldt & Barnes-Holmes, 2009) and is defined according to the presence of three core properties, termed mutual entailment, combinatorial entailment, and the transformation of stimulus function.[9]

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 functional relations that emerge between two or more mutually entailed stimuli. Thus, if A is bigger than as B and B is bigger than C, then humans will spontaneously derive that A is bigger than C as well as C is smaller than A. Finally, and perhaps most important within the context of the current work, once stimuli have been mutually or

[8] Although predicting the causal impact of mental events on subsequent behavior represents the *sine qua non* of mechanism as applied to psychology, this endeavor is not without its own difficulties. In particular, prediction 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 (see De Houwer, 2011 for a detailed discussion).

[9] 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 a learned behavior that emerges through a history of generalized operant responding (for a detailed treatment of how derived relational responding emerges see Törneke, 2010).

combinatorially related to one another, behavioral (psychological) functions may be transformed through that relation from one stimulus to another. Imagine for instance that an aversive function is established for stimulus A (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 never having been directly paired with a shock (e.g., Dougher, Augustson, Markham, Greenway, & Wulfert, 1994).

According to RFT the particular way in which stimuli are related, as well as psychological function(s) transformed through those relations, is determined by two forms of contextual control. On the one hand, humans can learn to relate stimuli in a vast number of arbitrarily applicable ways, including relations of 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 causality. Relational responding may also include deictic or perspective-taking relations (McHugh, Barnes-Holmes, & Barnes-Holmes, 2007). The contextual cues that establish and indicate which type of relational responding should be brought to bear on a particular situation are often referred to as "Crels". On the other hand, a wide range of psychological functions can be transformed through different stimulus relations, including discriminative (Dougher et al., 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), self-discrimination (Dymond & Barnes, 1996) and extinction functions (Dougher et al., 1994). The contextual cues that specify which functions should be transformed through such relations are termed "Cfuncs".

Over the past 20 years RFT has emerged as a coherent, parsimonious and progressive functional theory that has gained empirical momentum and stimulated a rapidly growing body of research (see Dymond & Roche, in press), clinical applications (Gaudiano, 2011; Rehfeldt & Barnes-Holmes, 2009), and conceptual controversy (Gross & Fox, 2009). Derived relational responding is argued to represent the fundamental behavioral unit underpinning a wide range of psychological phenomena including language and cognition (Hayes et al., 2001), intelligence (O'Toole & Barnes-Holmes, 2009; Cassidy, Roche, & Hayes, 2011) perspective taking and the self (McHugh & Stewart, 2012). As we shall see, the basic concept of derived stimulus relating in combination with the properties of relational complexity, coherence and derivation may also equip functional researchers with the conceptual tools necessary to understand, predict and influence implicit cognition.

6. A functional approach to the study of implicit cognition

While mutual, combinatorial entailment and transformation of function are the defining features of derived stimulus relating, two emergent properties of this type of responding are also worth noting in the context of the current paper. First, stimulus relations can vary in their *complexity* and be arranged along a continuum from low to high. 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. Not only can stimulus relations vary in their complexity but so too can the type and number of functions transformed according too those relations. For example, mutually or combinatorially entailed relations between stimuli may involve single functions being transformed based on a relation between one stimulus and another, whereas the relating of complex networks of relations to other networks may involve a vast array of stimulus functions being modified in accordance with those relations.

Second, relations can not only vary in their complexity but also in the degree to which they have been previously *derived* in the past. As noted above, derivation refers to the finding that once a set of relations between stimuli is directly trained, a number of additional untrained relations also emerge and allow for the transformation of function. To illustrate, consider a situation where a participant has just been trained to select B when given A and C when given B. Thereafter, and upon testing, a series of untrained relations are evident (e.g., selecting A when given C or C when given A). In this learning situation, we argue that the first instance in which the person derives the relation between A and C be defined as a "high derivation" response given that the history of deriving that particular response is minimal. Alternatively, imagine that the same person is then provided with an ever increasing number of opportunities to derive the relation between those same stimuli. Across each of these successive derivation opportunities the resulting response may come to be increasingly defined as involving "low" levels of derivation. Note that, according to RFT, derivation may well decline, with repeated instances, even when some form of programmed reinforcement is not provided for each derived response because derivation itself is "rewarded" by contacting increased relational coherence (see Hayes, Fox, Gifford, Wilson, Barnes-Holmes, & Healy, 2001, pp. 42–43). We will revisit this issue later in the paper.

In other words, and similar to relational complexity, the degree to which a response can be characterized as derived may also be arranged along a continuum from low to high. While complexity and derivation can be independently manipulated, they should not be conceptualized as mutually exclusive, but simply properties of relational responding that will be more or less evident in different contexts. In what follows, we draw upon these two response properties in order to explain (a) the convergent and divergent patterns of behavior observed on direct and indirect procedures, (b) why different classes of behavior are better at predicting certain outcomes than others and (c) to distinguish between the behaviors that are frequently divided into the domains referred to as implicit versus explicit. Towards this end, we first outline the core assumptions of this account, known as the Relational Elaboration and Coherence (REC) model, and then consider the benefits and challenges associated with explaining different patterns of behavior in this way.[10]

7. The relational elaboration and coherence (REC) model

A central premise of the REC model is that the behavioral outcome obtained on a measurement procedure will reflect an interaction between the individual's learning history with respect to the targeted relation(s) and the specific features of the context

[10] Note that the REC model is not distinct from RFT but hierarchically related to it. In other words, the relationship between RFT and the REC model is the same as the relationship between RFT and the model of analogical reasoning offered in Chapter 4, or the model of problem solving offered in Chapter 5, or the model of rule-governed behavior offered in Chapter 6 of the volume by Hayes et al. (2001). Indeed, if a second edition of the volume were to be published in the future it would contain a chapter on the REC model. This serves to underscore that RFT remains a work-in-progress and no volume or set of volumes should be seen as offering the final or complete account of human language and cognition. Rather, RFT is a constantly evolving analytic-abstractive position that feeds into, and is in turn fed by, its application to a diverse range of behavioral phenomena.

in which they are assessed. Stated more precisely, the behavioral process of arbitrarily applicable relational responding can be carved into different patterns of behavior that vary in their relative levels of complexity and derivation. Before continuing it should be noted that this functional account is characterized by two core assumptions that distinguish it from the mental models of implicit cognition discussed previously.

First, a mechanistic account of "automatic" and "controlled" behavior as the product of singular, dual or multiple mental processes operating exclusively or in interaction is substituted for a purely functional one. The REC model replaces the notion of causal forces or agency (mental constructs) with causal relations between behavior and environment. This reciprocal relation between behavior and environment can occur in a variety of ways, from direct learning contingencies established via respondent or operant learning, stimulus generalization and discrimination as well as those governed via arbitrarily applicable relational responding (e.g., instructions or inferences). Thus the probability of a particular response being emitted, whether that response is a button press on the IAT or the selection of a certain answer on a racial attitudes questionnaire, will either increase or decrease depending on the contingencies that governed similar responding in the past. In short, the mechanistic requirement for mental events to fill the spatial and temporal gaps between environment and behavior is replaced by a selectionist account (e.g., Darwinian natural selection) in which causation occurs over time and at a distance (see Chisea, 1994). In doing so, the notion of (implicit) cognition is emptied of any connotation of mediating forces or agency above and beyond the functional relation between environment and behavior.[11]

Second, cognitive researchers often conceptualize the (measurement) context as separate from the mental processes that are hypothesized to cause behavior and thus of secondary interest (e.g., Rydell & Gawronski, 2009). In direct contrast, and in-line with RFT, the REC model treats the behavior obtained on a measurement procedure as an on-going action that can only be understood with respect to the historical and current context in which it occurs. In other words, behavior *always* occurs within some context. When a change in behavior is defined as differences in the probability of a response due to interaction in and with the environment, the context takes on a central explanatory role. Thus for the REC model, the procedure – as a central feature of the current context – is not a passive set of instructions on how to generate a behavioral effect but rather an active explanatory factor in understanding that effect. More precisely, the degree to which different features of relational responding will be made evident to the researcher will depend on the specific constellation of conditions that characterize the context in which that behavior is measured (see Fig. 2).

At this point it seems important to note that previous work on the REC model suggested that responses 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). These preliminary sketches of the model focused on time as a means to distinguish the concept of BIRRs from EERRs,

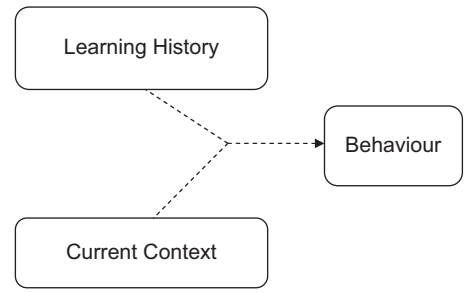

Fig. 2. Behavior causation according to the functional (contextual) approach. The individual's learning history with respect to the targeted relation interacts with features of current context in producing behavior on a particular measurement procedure.

but of course other properties would seem to be important in developing a more complete functional treatment of implicit cognition (see De Houwer & Moors, in press). Indeed, the foregoing arguments were never meant to be read as final or absolute functional definitions of automatic versus controlled behavior. In the following sections we focus on two additional properties of relational responding that may also be important in understanding implicit cognition; level of derivation and level of complexity (as will become clear, however, the parameter of time cannot be separated entirely from these two properties).

In keeping with a purely functional epistemology, the REC model arranges responding along these two probabilistic and graded continua, with behavior varying in *degrees* rather than in kind. In doing so, the notion of two dichotomous classes of "automatic" versus "controlled" behavior that are independent but mutually interactive is replaced with a more subtle discrimination between properties of arbitrarily applicable relational responding based on time, and more critically in the current article, levels of complexity and derivation. When viewed in this way, relational responding can be divided into four broad, overarching categories (see Fig. 3). These categories should not be rarefied as "static" or "permanent" but simply viewed as convenient visual illustrations for behaviors characterized by different levels of complexity and derivation. Likewise, the labels BIRRs and EERRs are not "static" or "permanent" things, constructs or mechanisms. They too are convenient labels that orientate the researcher towards a particular behavioral domain (implicit cognition), and in particular, properties of relational responding that seem important in understanding, predicting and influencing those behaviors. As will become clear, BIRRs often involve

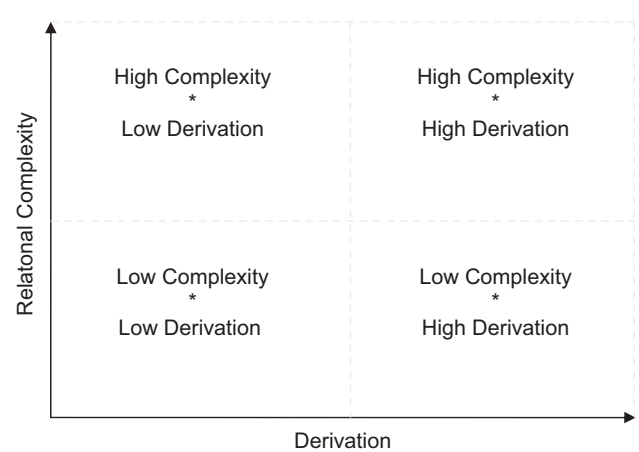

Fig. 3. Relational responding carved into four different categories as a function of the complexity and level of derivation that characterize the response.

[11] Note that while changes in behavior can refer to any somatic (e.g., button press on an indirect procedure), autonomic (e.g., salivation) or neural action (e.g., electrical impulses in the brain) that results from interactions with the current or historical context, functional researchers typically conceptualize behavior in terms of the *whole* organism. Nevertheless and with respect to implicit cognition, a number of (indirect) physiological procedures have been developed that focus on particular features of the behaving organism, such as electromyography (e.g., Roddy, Stewart, & Barnes-Holmes 2011), functional magnetic resonance imaging (Hart et al., 2000), event related potentials (Ito et al., 2004), and skin conductance (Dotsch & Wigboldus, 2008).

relatively lower levels of time, complexity and derivation, whereas EERRs involve relatively higher levels of these respective response properties. In the following sections we explore the possible features of each of these different response patterns in greater detail.

7.1. Low complexity low derivation responding

Given that relational responses, like all behaviors, unfold across time, the REC model proposes that (all things being equal) more complex responses may take additional time and be emitted with less accuracy relative to their less complex counterparts. To illustrate, consider the concept of nodal distance which refers to the number of nodes that link any two stimuli in a set of trained conditional relations. Interestingly, the time taken to respond in accordance with an equivalence relation increases and the accuracy of those responses decreases when the nodal distance within the equivalence class grows (Fields & Moss, 2007; Kennedy, 1991; Tomanari, Sidman, Rubio, & Dube, 2006; Wang, McHugh, & Whelan, 2012). Critically, however, when other relations above and beyond equivalence are involved, the complexity of a relation will be dictated not only by nodal distance but the number and type of relations involved (O'Hora, Roche, Barnes-Holmes, & Smeets, 2002; Steele & Hayes, 1991). Indeed, much work now indicates that as the number and type of relations increase the speed and accuracy of responding decreases relative to responses that are at lower levels of complexity (see Barnes-Holmes et al., 2005; Hayes & Bissett, 1998; Vitale et al., 2008; Reilly, Whelan, & Barnes-Holmes, 2005).

At the same time, the extent to which a response has been derived in the past will also influence its probability of being emitted quickly and accuracy in the future. For instance, the speed with which participants derive coordination, comparative and opposition relations becomes significantly faster with each successive opportunity to derive (O'Hora et al., 2002; Roche, Linehan, Ward, Dymond, & Rehfeldt, 2004; Steele & Hayes, 1991). Likewise, an overarching history of derivation facilitates the emergence of more accurate relational performances within and across stimulus sets (Bentall, Dickins, & Fox, 1993; Healy Barnes-Holmes, & Smeets, 2000; Roche et al., 2004; Sidman, 1994; Saunders & Green, 1999; Wang et al., 2012; Wulfert & Hayes, 1988). When taken together, it appears that the complexity of a relational response, as well as the degree to which it has been derived in the past, influences the probability that it will be emitted with speed and accuracy in the future. In particular, responses characterized by an extensive history of derivation and low levels of complexity appear to be emitted with relatively greater speed/accuracy than their more complex and less derived counterparts.

Insofar as the REC model is correct, then contexts that impose time and accuracy restrictions should increase the likelihood that responses will be low in both complexity and derivation. Over the past two decades, this methodological requirement for participants to emit low complexity relational responses under time pressure has been foundational in the development of many if not most indirect procedures (see Gawronski & De Houwer, in press). Indeed, while the original IAT and priming tasks – as well as their second generation variants – arrange the measurement context in a variety of ways, they each pair stimuli in close temporal or spatial proximity and require a low complexity (coordination) response to be emitted quickly. In such cases, speeded performances are used as a means to infer relatively simple relational responses between different classes of stimuli. What is important to note here is that by arranging the measurement context to primarily target coordination relations, the IAT and priming tasks are effectively restricted in the complexity of the learning history that they can capture. Put simply, while the individual may have a rich and complex history of relating with respect to the phenomena of interest, the vast majority of indirect procedures are intrinsically designed to target only the most rudimentary features of that history.[12]

Throughout much of the past 20 years mechanistic researchers have labeled low complexity (and we would argue low derivation) relational responses as "associative" and deployed them as evidence for the operation of mental associations in memory (Hughes et al., 2011). Importantly, however, the REC model argues that such responses are not associative at all—rather they are entirely relational in nature. According to this perspective, given a sufficient history of learning, and a measurement context designed to capture those relations, the behavioral effects obtained on indirect procedures should reflect other relational responses above and beyond coordination. In principle, BIRRs can involve any relationship between stimuli, such as opposition, hierarchy, spatial, temporal or comparative relations. Although the speed and accuracy of these responses will presumably vary in accordance with the levels of complexity and derivation of the targeted relation, there is no *a priori* reason why any type of relation should not be emitted quickly and accurately. One implication of viewing behavior in this way is that a non-associative indirect procedure is not only possible but quickly becomes necessary. Recently, a methodology has emerged from the functional literature that is capable of targeting relations that are low in complexity and derivation—the Implicit Relational Assessment Procedure (IRAP; see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010).

7.1.1. The IRAP

Similar to other indirect procedures, the IRAP arranges the measurement context in such a way as to place an individual's pre-experimentally established learning history into competition with a response contingency deemed inconsistent with that history of responding. Unlike any other indirect procedures developed however, the task is capable of targeting different types of stimulus relations (e.g. opposition, comparison) at differing levels of complexity (e.g., those that involve single stimuli or entire relational networks). To illustrate, consider Roddy, Stewart, and Barnes-Holmes' (2011) study examining brief and immediate relational responding towards overweight and average weight individuals. Participants received an IRAP on which a target stimulus (e.g., picture of an overweight or average weight person) was presented along with a label stimulus ("Good" or "Bad") and two relational response options ("Similar" and "Opposite"). During blocks of trials deemed consistent with a "pro-slim/anti-fat" learning history participants were required to emit the following relational responses: *Slim–Good–Same, Slim–Bad–Opposite, Fat–Good–Opposite, Fat–Bad–Same*. In contrast, blocks of trials deemed inconsistent with a "pro-slim/anti-fat" learning history required a reverse pattern of responding (i.e., *Slim–Bad–Same, Slim–Good–Opposite, Fat–Bad–Opposite, Fat–Good–Same*). In other words, 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 targets four separate stimulus relations independently of one another. In the above example, these stimulus relations were *Good–Slim*, *Good–Fat*, *Bad–Slim* and *Bad–Fat*. The difference in time taken to

[12] Although the IAT and priming tasks were primarily designed to target low complexity coordination relations, several researchers have obtained indirect evidence for other types of relational responding on these procedures, including opposition responding (e.g., "negation"; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000), deictic relating (Olson & Fazio, 2004), and evaluative relating (Förderer & Unkelbach, 2012).

respond on consistent relative to inconsistent trials – defined as the IRAP effect – is assumed to provide an index of the strength or probability of these various relations.

Interestingly, participants in the Roddy et al. study showed a significant IRAP effect for the *Good–Slim* and *Bad–Slim* relations, but no effect for the *Good–Fat* or *Bad–Fat* relations, indicating a positive relational bias towards average weight people but the absence of a negative bias towards their overweight counterparts. Furthermore, while performance on the IRAP was correlated with an analogous IAT, only the IRAP was correlated with participants' physiological activity registered via electromyography (EMG). Similar to the IRAP, the EMG measure indicated a pro-slim but neutral-fat response pattern. In other words, and consistent with the REC model, relational responses above and beyond coordination can be emitted with speed and precision. Across a growing number of studies and stimulus domains, BIRRs involving comparative (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), opposition (Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009) and distinction relations have been obtained (Nicholson & Barnes-Holmes, 2012). Moreover, IRAP performances have been found to correlate with those on the IAT and offer comparable if not better predictive validity when food related behaviors (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010), relational biases towards body-size (Roddy, Stewart, & Barnes-Holmes, 2010) and sexual orientation (Timmins, Barnes-Holmes, & Cullen, 2012) are assessed. Brief and immediate relational responding has also been found to correlate with a number of physiological (Roddy et al., 2011) and neurological indices (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008) in addition to other indirect procedures such as the Stroop task (Carpenter, Martinez, Vadhan, Barnes-Holmes, & Nunes, 2012) (for a detailed review see Hughes & Barnes-Holmes, in press).

Before continuing it should be noted that responses we suspect predominantly involve frames of coordination – such as the IAT and priming effects – will likely be emitted with relatively greater speed and accuracy than responses characterized by higher levels of relational complexity (e.g., IRAP effects that are likely based on comparison or distinction). More precisely, and given its centrality in natural language use, the generalized operant of coordination is likely at a higher probability relative to other relational operants and thus, in the absence of alternative controlling stimuli, may be emitted with a high degree of speed and accuracy (see Hayes et al., 2001). This is not say that coordination relations will necessarily always be the most "BIRR like" type of BIRR—rather like any functional interaction between the individual and his or her environment due consideration has to be given to the learning history and current context. For example, providing a long and well established history of opposition ("negation") relating across multiple contexts would likely result in such responses being emitted with greater speed and accuracy relative to their coordination counterparts (e.g., Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009; Kawakami et al., 2000; although see Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008).

To summarize, the REC model argues that performances on indirect procedures typically capture relational responses that are low in both complexity and derivation (see Fig. 4). The behavioral effect obtained from such tasks reflects an interaction between the (a) history of learning with respect to the targeted relation as well as (b) the constellation of features that characterize the measurement context. While a majority of indirect effects seem to be designed to capture coordination relations under time and accuracy restrictions, recent evidence suggests that other types of relational responses can also be emitted quickly and accurately.

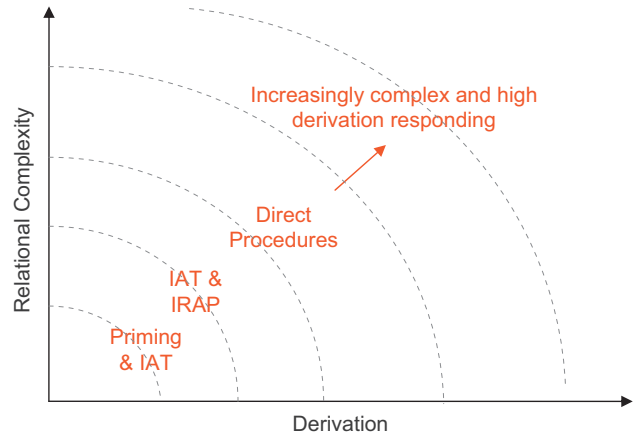

Fig. 4. Learning history brought to bear in a measurement procedure as a function of relational complexity and levels of derivation. The curved lines indicate that as levels of relational complexity and derivation changes, different measurement procedures will be differentially sensitive to targeting these relations.

7.2. Higher complexity low derivation

Although the previous section focused exclusively on responses that were low in their complexity and derivation, a more complicated picture emerges when these two factors interact. As we have seen, the REC model arranges relational responses along a complexity continuum, with the lower end of this scale occupied by BIRRs involving coordination (*Over-weight–Bad/True*), comparison (*Coke–Pepsi/Better–than*), distinction (*Arabic–Dangerous/False*) opposition (*Black–Bad/Opposite*), or, in principle, any other relational frame. As one travels further along this continuum the complexity of the related stimuli increases from single relations to those involving multiple relations and relational networks. These relations may be characterized by relatively higher levels of complexity yet still have been derived many times in the individual's learning history. Consequently, while these responses may be slower and less accurate compared to their low complexity/derivation counterparts, they should still be emitted with relative speed and precision.[13]

Consistent with this notion, Nicholson and Barnes-Holmes (2012) recently explored brief and immediate relational responding towards a variety of disgusting items, but using self-referential statements as target stimuli. During the IRAP participants were required to look at a picture of a disgust eliciting stimulus and then confirm or deny a statement that involves a complex deictic relational response (e.g., "*I am disgusted*", "*I need to look away*"). In effect, participants were asked to relate a picture to a relational network in terms of whether or not the network coordinated with a self-discriminated behavioral event. In total, two separate IRAPs were completed; one targeting so-called "propensity towards disgust" when confronted with revolting items (e.g., "*I am disgusted*") and a second assessing so-called "sensitivity" to the same stimuli (e.g., "*I need to look away*"). Not only did separate and significant IRAP effects emerge for these two types of relational responses, but they predicted entirely different behavioral outcomes on a set of behavioral approach tasks and self-report procedures. What is interesting here is that different relational responses (i.e., "*I am disgusted*" versus "*I need to look away*") gave rise to distinct outcomes on direct and indirect procedures—despite

[13] Within the implicit cognition literature, higher complexity/low derivation responding appears to have been mechanistically interpreted as the "contaminating" influence of higher order cognitive processes (i.e., propositions) on indirect procedures—which are typically viewed as targeting lower level processes such as associations (e.g., De Houwer, 2008).

the fact that participants were responding to the exact same set of pictorial stimuli on the two IRAPs.

Broadly similar findings have also been obtained when other higher complexity BIRRs have been subject to inquiry, such as self-referential statements about cocaine (e.g., "*With cocaine I am popular*") in a sample of drug users (Carpenter et al., 2012), self and body-size biases in a mixed group of patients with Anorexia Nervosa (Parling, Cernvall, Stewart, Barnes-Holmes, & Ghaderi, 2012), self-esteem in a prisoner population (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009) as well as emotional responding for high versus low depressed individuals (Hussey & Barnes-Holmes, in press; see also Nosek & Hansen, 2008). Although each of these studies seems to involve higher complexity relations than those discussed in the previous section, it is likely that these relations have been derived many times in the past given that they were emitted with relative speed and accuracy. Nevertheless, while these higher complexity relations may be, on average, at low derivation, relative differences in derivation levels across individuals may contribute to distinct outcomes on indirect procedures. For instance, it may be the case that "known groups" studies assessing the same level of relational complexity are in fact highlighting, amongst other factors, differences in the degree to which those responses have been derived in the past (e.g., Hussey & Barnes-Holmes, in press; Nicholson & Barnes-Holmes, in press).

As we continue further along the complexity continuum, the REC model predicts there will be some notional point beyond which the targeted relation will be too complex for reaction time based indirect procedures such as the IAT, IRAP and priming to capture. Although extensively deriving a relation may allow for the ensuing response to be emitted with relative speed and accuracy, this is nevertheless contingent on the complexity of the relation involved. Imagine, for example, a student of logic who has committed an extended syllogism to memory many times. After practicing this syllogism on one hundred occasions, she may be able to articulate the argument relatively quickly and accurately. However, the probability of the entire logical argument being emitted within a matter of milliseconds is relatively low (perhaps even impossible). Instead, this particular behavioral event will likely take place over an extended period of seconds (or even minutes) and thus fall outside the speeded response requirements implemented by many indirect procedures.

Stated more precisely then, when a response involves a complex interaction between multiple relational networks, increasing the level of derivation may not reduce the spatio-temporal properties of that response to the point that it can be effectively captured by time-based indirect procedures. Rather, an alternative methodology is needed that arranges the measurement context in such a way as to capture these extended and elaborated relational responses (EERRs). In particular, measurement contexts in which time is not a critical factor (e.g., semantic differential scales, likert scales, interviews, and focus groups) will likely allow for additional responses to occur toward the stimulus itself or toward the initial responses to that stimulus. More often than not, these direct procedures treat overt verbal reports as functionally equivalent to the EERR under investigation. Consider, for example, the question "*How do you feel about Brazilians?*" The ensuing response would presumably provide an index of the person's EERRs towards Brazilian people in general. Importantly, such procedures are not restricted to the assessment of ethnic stereotyping, but can target relational responding across a wide variety of content domains (e.g., preferences, opinions, beliefs, and values).

7.2.1. The problem with EERRs

Although the outcomes obtained from direct tasks potentially provide a more detailed insight into the individual's learning history relative to their indirect counterparts, they have long been recognized as susceptible to current contextual control in the form of demand characteristics and impression management (Holtgraves, 2004; Paulhus, 1989; although see Uziel, 2010). Put simply, people may recognize their private content but fail to report it, either because they do not endorse it, or because they are concerned about the social consequences of doing so. Indeed the ability for people to edit their overt responding in order to correspond with overarching rules implemented by the verbal community – or their own privately generated rules – serves to reduce the predictive validity of direct procedures in many domains (see Nosek et al., 2011). For instance, people are often reluctant to self-report engagement with illegal or socially taboo behaviors (Ames et al., 2007) or their affinity for violence when seeking treatment (Polaschek, Bell, Calvert, & Takarangi, 2010). Similarly, child sex offenders are hesitant to report their attraction to children yet differ in their brief and immediate relational responding towards "children as sexual" compared to non-offenders (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009) and violent offenders with no sexual crimes towards children (Brown, Gray, & Snowden, 2009).

The REC model suggests that rule governed behavior may provide one explanation for why outcomes on direct procedures often lack predictive validity as well as diverge from indirect procedures assessing the same phenomena. According to this perspective, rules modify or establish the psychological functions of events via a transformation of functions that typically involve coordination, comparison, temporality and/or causality. They constitute a class of EERRs that can be established by the individual themselves or the wider verbal community and specify a contingency between antecedent, behavioral and consequential events (see Ju & Hayes, 2008; Törneke, 2010; Whelan, Barnes-Holmes, & Dymond, 2006). Importantly, the behavior or consequences specified by the rule need not have been directly contacted by the individual in the past. What is necessary is a history of relational responding, coupled with previous contact with the consequences of rule following. Imagine, for example, you are instructed that "*public declarations that black people are bad are typically punished by society in a variety of ways, such as being verbally reprimanded, socially ostracized or imprisoned*". This rule places you in contact with a relational network that transforms the functions of the stimuli within that network. In the above example, overt negative evaluations of black people are related temporally and causally with a particular class of aversive stimuli. Terms such as "punished by society" alter the functions of behavior (i.e., making negative public statements about black people), and the transformation of stimulus functions provides these actions and contexts with some of the features of the specified consequences (i.e., avoidance). As the person responds in accordance with the rule, the coordination between the rule and "the relations sustained among the specific events when the rule is followed, provides an on-going source of control over behavior" (Hayes et al., 2001, p. 107). Put simply, when a rule is followed, events in the environment participating in that relational network acquire verbal functions as specified by the rule. Thus if an individual who has historically emitted discriminatory remarks about black people stops doing so after encountering the above statement, it would seem that their behavior is meaningfully rule-governed.

In regulating the behavior of its members, the verbal community establishes and enforces a wide range of rules for appropriate conduct, and this is often the case for classes of stimuli that are relevant to "socially-sensitive" domains such as race, sexual orientation, religion and politics. Importantly, however, these rules may or may not correspond to the individual's own history of responding. It is also worth noting that people can understand these rules and yet fail to follow them. Consider, for example, a white woman with a history of relating black people with negative stimuli who is asked to complete a questionnaire

assessing her behavioral tendencies towards this particular racial group. On the one hand, the women may disregard her prior history of relating as the basis on which to respond and follow an overarching rule operating within that context (e.g., report a lack of racial bias). In this speculative example, verbal regulation of behavior will likely result in a divergence between responses on the questionnaire and her behavior towards black individuals in general. On the other hand, the woman may "disregard the rule" in contexts where socially mediated punishment seems unlikely (e.g. when talking to a friend who she knows shares her racial biases) and respond in accordance with her prior history of relating black people with negative stimuli. Within experimental situations, contextual factors such as guarantees and evidence that self-reports are anonymous, assessing the relational response unobtrusively, or hiding the true intent of the question may lead to responding that remains unaffected by rules pertaining to the social-acceptability of expressing negative opinions about the stimuli, and thus an increase in correspondence between direct measures and future behavior may be observed (e.g., Devine, 1989; Fazio et al., 1995). Likewise, soliciting responses from participants who believe that the experimenter can readily identify deception may also serve to increase correspondence across measures (e.g., "Bogus pipe-line"; Roese & Jamieson, 1993).[14]

To summarize, when a history of reinforcement has previously been contacted for responding in-line with socially- and self-generated rules (e.g., "In general, report positive evaluations of black people"), these rules (EERRs) may influence the probability with which functionally similar EERRs are publically emitted in similar contexts (e.g., indicate a positive racial bias on a self-report questionnaire). Given that direct procedures typically arrange the measurement context in a non-time pressured fashion and specify the relational response under investigation, the individual has sufficient time to emit a BIRR (e.g., a negative evaluation of a black person), an EERR ("In general, report positive evaluations of black people"), followed by a response to the BIRR and EERR (e.g., "Ignore initial reaction and tick positive response box on a questionnaire). Consequently, it becomes difficult, although not impossible, to "carve out" one response from another in the verbal stream of behavior. Doing so requires that specific contextual manipulations be implemented during the experiment that increase the likelihood that responding remains unaffected by socially- and self-generated rules about the stimuli under investigation. Finally, and before proceeding, it is important to realize that rules are functional units like all of the behaviors discussed thus far and do not constitute anything new at the level of behavioral process. They are simply examples of the interaction between complex relational networks and the transformation of function through those networks (i.e., relationally complex responses).

7.3. Low complexity higher derivation

So far we have focused on relational responses that vary largely in their respective levels of complexity. Yet the REC model also draws attention to situations in which a response is at low complexity but varies in its level of derivation. For instance, low complexity responses at the upper boundaries of the derivation continuum are likely to have been emitted only a small number of times (e.g., the very first testing trials on equivalence or stimulus relating tasks). As noted above, these responses may initially be slow and prone to error. However, as the opportunity to derive the relation increases – both within and across successive training and testing contexts – so too will the relative speed and accuracy of the emitted response (e.g., O'Hora et al., 2002; Roche et al., 2004).

Throughout much of the past 20 years functional researchers have exploited relative differences in derivation levels as a means to probe the strength or probability of pre-existing stimulus relations. In an early study in this vein, Watt, Keenan, Barnes and Cairns (1991) attempted to establish a three member equivalence class using Northern Irish and English participants. During the task a series of conditional discriminations were successfully trained in which Catholic family names were matched to non-sense syllables and those same non-sense syllables were then matched to Protestant symbols. Interestingly, when the researchers then tested for equivalence responding (i.e., matching Catholic names directly to Protestant symbols) many of the Northern Irish (but not English) participants failed the task. In other words, it appears that pre-existing verbal relations with a high probability of having been derived many times in the Northern Irish community disrupted the formation of low complexity relations that were inconsistent with that history of learning. This finding equipped researchers with a novel means to test the strength of pre-existing relational repertoires concerning clinical anxiety (Leslie et al., 1993) and social discrimination (Dixon, Rehfeldt, Zlomke, & Robinson, 2006) not to mention self-knowledge (Merwin & Wilson, 2005) and child sex abuse (McGlinchey, Keenan, & Dillenburger, 2000). More generally, this strategy of pitting relations with the same level of complexity (but different levels of derivation) against one another is foundational to many indirect procedures advanced thus far. Tasks such as the IAT and IRAP arrange the measurement context so that one response contingency (responding in accordance with pre-experimentally established stimulus relations) is put into competition with another (responding in-line with laboratory induced high derivation relations that are inconsistent with that history of responding) so that relational response strength can be ascertained.

As we continue further along this continuum from higher to lower derivation, the REC model suggests that responses should increasingly be emitted with speed and precision. Consider for example newly established histories of relating generated within an experimental context via instructions, inferences or statements (e.g., Gast & De Houwer, 2012; Rydell & Gawronski, 2009; Smith, De Houwer, & Nosek, in press), evaluative conditioning (Hofmann et al., 2010), minimal groups manipulations (Paladino & Castelli, 2008) or conditional discrimination tasks (Hughes & Barnes-Holmes, 2011). More often than not, these laboratory preparations involve a single training session aimed at generating low complexity relations towards a range of novel stimuli such as fictitious individuals, social groups, consumer brand products and non-sense words. Although a relatively short history of derivation in combination with low complexity appears to result in quick and accurate responding, it is important to note that these relations have been derived far less often than many others in the individual's verbal repertoire. Consequently, it may be the case that relations that are low in complexity but high in derivation are more malleable to change relative to those that are characterized by an extended and well-established history of derivation. For instance, findings related to US revaluation (Walther, Gawronski, Blank & Langer, 2009), approach-avoidance training (Woud, Becker & Rinck, 2008), persuasion (Smith et al., in press) and descriptive information (Rydell & McConnell, 2006) reveal that experimentally induced histories of relating can be generated as well as modified within a single experimental session.

7.4. High complexity high derivation

Finally, at the furthest corners of the coordinate system reside responses that are exceptionally high in both complexity and derivation. Such cases may involve deriving the relation between

[14] Rules are not exclusively restricted to "socially-sensitive" domains but may also govern behaviour in other non-socially sensitive areas. For instance, a participant may respond in accordance with the following "demand compliant" rule: "*I think the experimenter wants to find X. Ergo, I will respond in-line with X*".

a set of inter-related relational networks a small number of times or even for the very first time. Behaviors such as insight, creativity, abstract mathematics (e.g., Ninness et al., 2006), complex problem-solving (Stewart, Kelly, McHugh, Barnes-Holmes & O'Hora, submitted for publication), perspective taking (McHugh & Stewart, 2012) and analogical reasoning (Stewart & Barnes-Holmes, 2004) may qualify as examples that broadly fall within this area. Although a detailed treatment of these responses is beyond the scope of a paper on implicit cognition, it is important to realize that these behaviors are still relational and thus bound by the same principles that govern all other types of derived stimulus relating discussed thus far. To our knowledge, this particular area of the coordinate system is relatively uncharted territory as far as an experimental RFT analysis goes, particularly with regard to creativity, insight, and problem-solving.

Although each of the above phenomena may initially qualify as instances of high complexity/derivation responding, it is important to realize that as their respective levels of derivation increases, their relative position along our coordinate system will change. Thus while analogical relations, and the transformations of function that occur based on those relations, may be in principle highly complex the very first time they are derived, subsequent responses to those same relations may be emitted with greater speed and accuracy across successive derivation opportunities (e.g., Barnes-Holmes et al., 2005). The same goes for perspective taking (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004). Although speculative, it may be the case that as derivation moves from high to low so will the complexity of the ensuing response. To illustrate, consider the following relational network: *moderate pain is the same as jiget and jiget is less than werte*. In the first instance participants may derive elements of the network by initially considering each of the mutually entailed relations ("jiget is the same as moderate pain and werte is more than jiget") prior to relating those stimuli in a combinatorial fashion ("therefore werte is extremely painful). In this case, the derived response will likely take time to emit and may initially be low in accuracy. As the number of derivation opportunities increases, however, participants may respond directly to werte as extremely painful. Whereas the complexity of the original relation does not change topographically, functionally speaking, responding to werte as extremely painful following an extensive history of derivation may involve a less complex relational response compared to the first time it was emitted. Put differently, a single stimulus from a complex relation may come to capture the functions of the entire network without the need for the individual to derive all elements of that relation. For example, an individual who is familiar with the story of the "Good Samaritan" may respond to these two words in an IRAP as functionally equivalent to the entire story as it appears in the Bible.

7.5. Conclusion

The current paper proposes a new theoretical model for the study of implicit cognition that is based on two relatively simple notions. The first is that complex human behavior (such as explicit and implicit cognition) reflects the learned and contextually controlled ability to arbitrarily relate events to one another. The second is that these relational responses can be arranged along two probabilistic and graded continua as a function of their relative complexity and history of derivation. In other words, the REC model attempts to provide a theoretical integration of the heterogeneous findings obtained from direct and indirect tasks using two properties of derived stimulus relating; complexity and levels of derivation (and their temporal parameters). While a wide range of topographically distinct behaviors are subsumed under one theoretical umbrella, no radically different processes or constructs are called upon to accommodate the movement from simple to complex behavior—just the idea that people can arbitrarily relate stimuli to one another in increasingly complex ways. From punctuate priming and IAT effects to multiple stimulus relations captured by the IRAP, right through to responses on self-report questionnaires, extended narratives, analogical reasoning, problem solving, creativity, self and perspective-taking, we argue that it is all relational. Moreover, we propose that different measurement procedures will be more or less sensitive to brief and immediate versus extended and elaborated responses depending on the manner in which they arrange the measurement context.

8. BIRRs and EERRs: the role of relational coherence, predictive validity and awareness

Over and above these various patterns of relational responding, the REC model makes specific assumptions about the interaction between BIRRs and EERRs and their predictive relationship with other classes of behavior. In the following sections we articulate when and why different relational responses will either converge or diverge from one another and predict specific types of behavior. Thereafter, we briefly consider the "awareness" issue that has occupied a great deal of theoretical and empirical attention within this research area.

8.1. Relational coherence

A well-documented finding in the implicit cognition literature is the possibility for BIRRs to either conflict or correspond with EERRs towards the same stimuli. For example, several studies have demonstrated changes in EERRs but not BIRRs (Gawronski & Strack, 2004; Rydell & McConnell, 2006), BIRRs but not EERRs (Dasgupta & Greenwald, 2001) while still others have reported corresponding changes in both BIRRs and EERRs (Gawronski, Walther, & Blank, 2005; Richeson & Nussbaum, 2004). In explaining these various outcomes the REC model appeals to a third property of derived relational responding (i.e., relational coherence) that has not been discussed so far in the current article. Specifically, a relation is said to cohere when all the individual elements relate to each other in a manner that is consistent with the reinforcement history typically provided by the verbal community for such responding. Although non-human animals show evidence for behavioral consistency based on direct contingency learning and the formal properties of the related stimuli, humans alone appear to search for and create consistency between and among derived stimulus relations involving arbitrary stimuli. To illustrate, imagine you are presented with the following statement "10 dollars is more than five dollars and five dollars is more than one dollar therefore one dollar is more than ten dollars". It is likely that you would recognize the incoherent nature of this simple relational network and question its veracity. As noted earlier, once a history of relational responding is in place, coherence quickly becomes a type of conditioned reinforcer for future derived stimulus relating (see Barnes, Hegarty, & Smeets, 1997; Hayes et al., 2001; Quiñones, 2008; Vitale et al., 2008).

According to the REC model, this search for and application of relational coherence applies equally to BIRRs and EERRs. Imagine that you are a "chauvinistic" male with an extreme fear of flying who is boarding an airplane. Upon entering the cabin you notice that both the pilot and first-officer are female and immediately experience momentary surprise as well as the temptation to cancel your flight and reschedule for an alternative date. On the one hand, this initial BIRR may conflict with other, subsequent relational responses that follow (e.g., "*Anyone can be a pilot regardless of gender*"). Given a history of responding coherently

in the past, you may engage in additional relational activity to reduce or resolve this incoherence such as responding to your initial response as "wrong". Within an experimental context, this should lead to discrepant outcomes on direct and indirect procedures. At the same time, you could consider your initial response as "correct", but instead choose to respond in-line with prevailing rules implemented by the wider verbal community (e.g., "*Negative evaluations based on gender are inappropriate in modern society*"), or even self-generated rules (e.g., "*I will look like a chauvinistic pig if I refuse to fly with a female air crew*"). In such cases, overt responding may be strategically altered prior to being emitted to better concord with prevailing social norms or the presumed expectations of the researcher. Once again, within an experimental context, this should lead to a dissociative pattern of outcomes on direct and indirect procedures.

On the other hand, it is also possible for BIRRs to be consistent with subsequently emitted EERRs and as a result lead to convergence between direct and indirect procedures. In the example above, the individual may conclude that female pilots would be less capable of dealing with an emergency situation relative to their male counterparts, which would cohere with their initial relational response (and result in refusing to board the aircraft). In an experimental context, such behavior should lead to corresponding outcomes across both direct and indirect procedures. Although the foregoing analysis suggests that BIRRs generally precede and are distinct from EERRs, it is important to realize that they are both relational responses, and thus will interact in a dynamic fashion. Indeed, an EERR may well generate a BIRR in a given stream of verbal behavior and vice versa. Applied to the current example, the individual may emit the EERR "*It is wrong to discriminate on the basis of gender*". This may in turn result in a brief negative self-evaluative response (e.g., "*I am a chauvinistic pig*") which in turn produces another EERR ("*Okay, calm down, relax, take some deep breaths, get on the plane and watch a movie*"). While this relationship between BIRRs and EERRs is potentially informative, we are not suggesting that the former provides the causal explanation for the latter. On the contrary, these behavior–behavior relations are always traced back to the point at which prediction and influence are in principle directly possible (i.e., manipulable contextual variables; see Hayes & Brownstein, 1986).

8.2. Prediction of behavior

In addition to the above mentioned patterns of convergence and divergence, the outcomes obtained from indirect procedures often predict a number of behaviors that their direct counterparts fail to accommodate (and vice versa). In interracial interactions, for example, indirect effects predict the rate of blinking and eye contact with a member of another racial group while performance on direct procedures better predict verbal biases in evaluating a Black relative to a White interviewer (e.g., Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; see also Fazio et al., 1995; McConnell & Leibold, 2001). In a similar vein, Richetin, Perugini, Prestwich and O'Gorman (2007) found that the spontaneous selection of either fruit or a snack was predicted by IAT performance over and above the prediction provided by direct measures for behavioral food choice.

In explaining the predictive relationships between direct/indirect effects and future behavior, the REC model draws attention to the fact that BIRRs and EERRs are functionally defined classes of behavior that are not restricted to any particular topography. As noted above, BIRRs can – in principle – not only take the form of a button press on an indirect behavioral task, but increased rates of blinking, eye contact, and "spontaneous" selection of stimuli given an appropriate history of learning. Likewise, EERRs are not restricted to written or oral responses on a questionnaire but can involve the decision to invite job applicants for an interview (Agerström & Rooth, 2011) or supporting exclusionary immigration policy proposals (Perez, 2010). From this perspective, the ability for one type of response (e.g., IAT performance) to predict another (amount of eye contact) is understood functionally as an instance of behavior–behavior relations. Thus when we ask "What role does automatic and controlled thinking play in predicting different types of behavior?" we are actually asking about the nature of a behavior–behavior relation. In doing so, the question of how mental mediators account for differential patterns of behavior is immediately discarded in favor of an alternative question: what are the contingencies that will influence the probability of one class of responses predicting a second class of responses?

Unsurprisingly, the functional researcher once again looks to the current and historical context for an answer. Within the wider literature, the predictive relationship between BIRRs/EERRs and future behavior appears to be influenced by the specific stimuli being related (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), features of the context that influence the probability that behavior will be governed to a greater or lesser extent by BIRRs or EERRs (Marsh, Johnson, & Scott-Sheldon, 2001), instructions to self-discriminate brief and immediate responding towards the stimuli of interest (Scarabis, Florack, & Gosejohann, 2006), the time available to emit a response (Friese, Wänke, & Plessner, 2006), the mood of the respondent (Hermsen, Holland, & van Knippenberg, 2006), as well as the degree to which self or socially generated rules are deployed (Olson & Fazio, 2004). Although other types of predictive relationships exist between some BIRRs/EERRs and other BIRRs/EERRs (for a detailed treatment, see Nosek et al., 2011; Perugini et al., 2010) a REC explanation of these findings remains firmly rooted at the contextual level—irrespective of the complexity involved. While a number of studies have attempted to account for the predictive relationship between BIRRs and other types of behavior (Barnes-Holmes et al., 2010; Nicholson & Barnes-Holmes, 2012; Roddy et al., 2010; 2011), only simple associative or additive patterns have been examined. A more detailed understanding of the relationship between BIRRs/EERRs and future behavior awaits a functional analysis of increasingly complex patterns of behavior–behavior relations than is currently available (e.g., double-dissociation/additive; interactive/multiplicative or partial dissociation patterns; see Perugini et al., 2010).

8.3. The "awareness" issue

As we have seen, people may be "unwilling" to report their private relational content due to self or socially generated rules. However, it is equally possible that they are not "aware" of their private content to begin with or how it relates to their public behavior. Indeed, research shows time and again that individuals are often limited in their ability to self-report their brief and immediate relational activity—even in situations where they are motivated to do so (Nisbett & Wilson, 1977; Wilson, 2009). As one of the "four horsemen of automaticity" (Bargh, 1994), a number of mechanistic researchers have treated "awareness" as a central or even defining property of implicit cognition (e.g., Greenwald & Banaji, 1995; Nosek et al., 2011; although see Gawronski & Bodenhausen, 2011).

Critically, awareness is not a technical term from a functional point of view and we do not intend to make it one. Nevertheless, in some meaningful sense of the word, derived stimulus relating – and by implication – implicit cognition can occur with or without "awareness". The REC model argues when a person makes a self-report concerning their own brief and immediate relational responding that BIRR is likely participating in a frame of

coordination with deictic relational responding (e.g., "*I just had a negative reaction to that black person*"). When treated in this way, awareness is emptied of any meta-physical or mental conations and instead is viewed as a learned, self-discrimination behavior that develops over time as a history of relating relations to other relations – and in particular deictic relations – becomes increasingly well established (see Dymond & Barnes, 1996, 1997; Hayes et al., 2001). As always, this behavior is under contextual control. Thus when the parameters of the context permit or require the individual to respond to their on-going relational activity with deictic relations about that activity, an overt report of BIRRing should be available. Though ultimately an empirical question, we expect that a lack of coordination between on-going relational activity and deictic relations about those relations is most likely to occur with exceptionally low complexity/derivation responses that have been relatively well practiced in the past. At the same time we would expect that awareness or self-discrimination would increasingly occur as the complexity of the relation becomes greater or where a history of deriving that relation is minimal. Finally, it is important that the self-discrimination of a brief and immediate relational response is separated from self-discrimination of the history of learning that gave rise to that response. In the latter case, the individual may respond on the basis of their BIRR but may not be able to articulate why they had that BIRR in the first place.

9. Implications of the REC model

The current paper has now introduced three new conceptual tools (relational coherence, complexity and derivation) as well as a methodological one (the IRAP) that functional researchers can use when exploring new intellectual islands. These concepts are non-mediating or mentalistic but simply to refer to properties of the same behavioral process (arbitrarily applicable relational responding) that become more or less prevalent in different (measurement) contexts. Like all definitions, the parsing of relational responses based on time, complexity, or derivation is a matter of convention; not fixed or absolute but rather flexible to further modification in-line with empirical findings. The "truth" or value of the REC model will be a function of how useful it is in providing a more sophisticated understanding, defined in terms of achieving greater prediction and influence over private and public behaviors (with precision, scope and depth) than is currently available. Although this account requires in-depth empirical scrutiny, we believe that it can allow functional researchers to achieve the above goals in a number of novel ways.

9.1. General conceptual issues

By treating implicit cognition as relational responses that can vary in their complexity and derivation, the REC model orientates empirical attention away from the traditional notion of "automatic" and "controlled" behaviors and towards a view of implicit cognition as probabilistic and graded. This account offers a coherent framework that can accommodate a spectrum of behavioral outcomes, from those that take place within an order of milliseconds all the way to those that unfold across extended periods of time. One potential advantage of this position is that it may account for outcomes that do not fall neatly into "automatic" or "controlled" categories (i.e., reside in the middle of the complexity-derivation continuum). For mechanistic researchers a range of effects are located in the "fuzzy boundary" between automatic and controlled behaviors such as "semi-automatic negations" (Deutsch et al., 2009), propositional effects on indirect procedures (De Houwer, 2006; Gast & De Houwer, 2012; Gregg et al., 2006), and the ability to control, fake, or modify performance on indirect tasks (Bar-Anan & Nosek, 2012; Fiedler & Bluemke, 2005). The REC model seems to provide a rationale for these findings by emphasizing the utility of moving away from binary distinctions and towards probabilistic, graded alternatives (see De Houwer & Moors, in press; Nosek et al., 2011 for similar suggestions from the mechanistic literature). Although we use terms such as BIRRs and EERRs throughout the latter sections of the current paper, they should not be considered binary or dichotomous but rather convenient ways of talking about responses that differ in their *relative* levels of complexity and derivation.

Our account also provides functional researchers with a nomenclature that is not imported from either the lay community or alternative intellectual traditions. Each of the terms discussed above are directly rooted in a bottom-up functional theory that coherently connects basic and complex concepts. One of the advantages of doing so is a reduced tendency for different levels of analysis to be conflated with one another (De Houwer, 2011). It should now be apparent that while mechanistic and functional researchers are interested in similar behavioral events, BIRRs are not "implicit", "unconscious" "spontaneous", "associative" "impulsive" or "automatic". Likewise, EERRs are not "explicit", "conscious", "deliberate", "propositional", "reflexive" or "controlled". Instead the notions of complexity and derivation open the concepts of BIRRs and EERRs to a more precise and philosophically consistent experimental analysis. Questions surrounding the formation, change and interaction of BIRRs and EERRs, their malleability in the face of historical and current contextual variation, convergence and divergence, relationship with relational coherence, as well as their role in clinical, social, health and other psychological domains can be asked and answered without appeal to explanatory concepts that are incoherent with the basic theory (RFT) or the assumptions and goals that underpin contextual behavioral science.

9.2. Levels of derivation rather than verbal versus non-verbal

The focus on levels of derivation within the REC model serves to highlight an important issue that has not been fully appreciated within the general RFT literature. Specifically, 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, this distinction may be problematic for several reasons. First, and foremost, an experimental demonstration of "non-verbal" behavior presupposes that the individual can set aside his or her history of arbitrarily applicable relational responding that is brought to bear in any given context. Consider a situation where a verbally trained individual is asked to complete a respondent learning task where a non-sense word (conditioned stimulus; CS) is repeatedly paired with an electric shock (unconditioned stimulus; US) across one hundred separate trials. Following training, the previously neutral CS acquires an aversive function. 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. Now imagine a second scenario whereby a novel equivalence class is established consisting of A, B and C. The first time the individual derives the relation between A and C could be considered as a "highly verbal" behavior. However, across successive opportunities to derive, the response may be said to become less derived and more a product of the direct contingencies contacted within that context (e.g., direct pairing of A with C, reinforcement contacted for responding coherently with previous training, or progressing from training to testing). Or to

put it another way, as one derives more and more, behavior becomes less and less "verbal".

From an RFT perspective, however, all behavior based on a history of arbitrarily applicable relational responding is, by definition, verbal—irrespective of whether it is increasingly characterized by less and less derivation (see Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000, pp. 72–75). With respect to the current paper, although low complexity/derivation responses (BIRRs) 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. To illustrate, 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 is at low derivation 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.

With the above in mind, we would remind the reader that RFT is an intensely functional and pragmatic theory of language and cognition. Specifically, once a protracted history of arbitrarily applicable relational responding is in place, it seems useful to define everything that an individual does thereafter as involving verbal processes. Indeed, it is difficult to imagine a behavior that is not impacted in some way, shape or form by this 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. Thus we believe that the verbal/non-verbal distinction should be jettisoned in favor of one that differentiates behaviors based on their respective levels of derivation; from situations where derivation is minimal and contact with direct contingencies maximal (e.g., "mindful" activities such as meditating, "getting lost in the music") to those where derivation is maximal and contact with direct contingencies is minimal (e.g., solving a complex mathematical or logical problem for the first time).[15] Note that this is not an ontological argument—merely a pragmatic one. For example, researchers can explore the extent to which a response varies in speed and accuracy or is strengthened or extinguished as a function of derivation without taking on the additional burden of proving that it is also a verbal or non-verbal act as well.

9.3. Limitations and future directions

We believe that the REC model in conjunction with the IRAP provides functional researchers with an opportunity to participate fully in the study of implicit cognition alongside our contemporaries in social and cognitive psychology. That said, a number of empirical questions require our attention at this time. First and foremost, although a limited set of RFT studies have sought to manipulate relational complexity and derivation, a detailed experimental analysis of the interplay between these two factors in the production of different types of relational responses (i.e., BIRRs and EERRs) is clearly needed. This applies equally to relational coherence and "awareness" as a self-discriminatory behavior—both as topics in and of themselves as well as their interaction with the above two factors. To our knowledge no research has directly explored the intersection between relational coherence, complexity and derivation with respect to implicit cognition. Likewise, although "self-awareness" has been explored elsewhere, its treatment in relation to performances on direct and indirect procedures is yet to be functionally articulated. Embarking on the above research agendas would refine our understanding of when and why BIRRs and EERRs independently or interactively predict different classes of behavior.

Second, while a small number of studies have provided experimental evidence for the emergence of BIRRs (e.g., Hughes & Barnes-Holmes, 2011; O'Reilly, Roche, Ruiz, Tyndall, & Gavin, 2012; O'Toole, Barnes-Holmes, & Smyth, 2007) a systematic exploration of the learning histories and current contextual variables critical to establishing, maintaining, and changing this class of behaviors is clearly needed. Although IRAP and related research provide support for the relational nature of implicit cognition, this work has almost exclusively focused on manipulating pre-existing BIRRs and EERRs across a wide variety of domains (Hughes & Barnes-Holmes, in press). Subjecting the environmental regularities that influence these different classes of behavior to an experimental analysis would serve to provide a more rigorous test of the assumptions outlined throughout the current paper. Moreover, the degree to which high complexity/derivation responses (EERRs) can govern their less complex/derived counterparts (BIRRs) is also in need of exploration.

Third, the development of BIRRs and EERRs across the lifespan also needs to be addressed. Of the RFT research conducted with infants, children and adolescents, very little has investigated the development and change of implicit cognition over time. Rather functional researchers interested in this topic have exclusively focused their attention on adult populations who already have an extensive history of relating. Given the operant nature of relational responding, we would expect BIRRs and EERRs to demonstrate four important properties across all stages of the lifespan once a history of arbitrarily applicable relational responding is in place: (a) they should develop, (b) be flexible and amenable to change, (c) come under stimulus control, and (d) be controlled by their consequences. Although supportive research has been provided on all four points for EERRs in adults as well as children (Barnes-Holmes & Barnes-Holmes, 2000; Luciano, Gómez-Becerra, & Rodríguez-Valverde, 2007), little work has focused on examining the same properties with respect to BIRRs (although see Hughes & Barnes-Holmes, 2011; O'Toole et al., 2007 for preliminary work in this vein). Although pragmatic considerations can interfere with the collection of data in infants, we argue that a developmental understanding of implicit cognition is certainly worth the effort (see Banaji & Heiphetz, 2010, for a discussion of related work in this area).

[15] These conceptual concerns are particularly salient for research with adults who have an extensive history of relational responding. When the subject of interest is non-human research or human infants without a history of appropriate relational responding, the verbal/non-verbal distinction seems to be a more acceptable and useful one. For instance, given the "Pretty Polly" example above, one could imagine that a human infant without an extensive history of arbitrarily applicable relational responding, but who had received an extensive training in echoing "Pretty Polly", would simply continue to engage in that echoic response when asked who is the opposite of "Pretty Polly".

A fourth direction for future work could examine BIRRing and EERRing in physiological and neurological domains. Throughout the current paper we have situated our analysis in terms of the whole organism operating with and in a given context. In doing so, we have intentionally downplayed any emphasis on different levels of analysis (e.g., genetic and neurological) and their role in relational responding—both fast and slow. Although we wholeheartedly support a functional investigation of implicit cognition at these levels, it is important that causality continues to be framed in terms of events external to the behavior of the individual and not in terms of elements of the behavioral system that needs to be explained (see Hayes & Brownstein, 1986 for several reasons why). Nevertheless, when behavioral genetics and neuroscience are contextually situated, our understanding of environment–behavior interactions can only be improved.

Finally, it is important to note that the REC model is an overarching account of behavior that is not restricted to any single behavioral domain (e.g., IRAP effects). The notion of BIRRs and EERRs may prove useful for understanding, predicting and influencing a host of clinical, health, consumer, forensic, close-relationships, and organizational, political, sport and social behaviors. Our intention was to frame our analysis in a way that could lead directly to empirical tests and extensions. Time will tell whether we are successful or not. In the absence of data, a parsimonious and overarching functional theory such as RFT can be used to explain many behavioral phenomena in a reasonably coherent and plausible fashion. Fortunately however, actual data serve to exert a selective effect upon such theorizing and we are exquisitely sensitive to this point. In closing, therefore, we encourage functional researchers to take, use, modify and develop RFT and the REC model in their own empirical voyages to uncharted waters and their exploration of foreign intellectual islands.

10. Epilogue

Our treatment of the *a priori* assumptions and analytic strategies adopted by mechanistic and functional researchers was driven by two complementary goals. First, by making explicit the scientific beliefs, basic units, as well as values that different researchers adhere to, the rationale for past and future developments within this research area can be better understood. Indeed, our analysis makes it clear that neither the mechanistic nor the functional approaches should be viewed as *the* conceptualisation of implicit cognition, but instead as co-existing conceptualisations of implicit cognition. Moreover, any evaluation of theories, models or procedures from either tradition that fails to consider the overarching goals, values or assumptions upon which that program of research is based is either flawed or simply incomplete.

Our second goal was to illuminate the subtle and seductive dangers of conflating the cognitive (mechanistic) and functional (contextual) approaches to psychological science. This is a particularly salient issue in the study of implicit cognition, which constitutes a rich melting pot of methodologies, terminology and theories from either tradition. Unlike clothing items, candy or food combinations, "mixing and matching" these frameworks should be avoided at all costs given their different viewpoints on the goals of science, explanatory concepts and how evidence should be treated. Although these points of departure may seem irreconcilable and suggest that fruitful dialog is improbable or even undesirable, we believe that such a position is detrimental for all concerned. Rather than subject one another to derision and scorn we believe that researchers from both traditions would profit from the recently proposed functional-cognitive framework advanced by De Houwer (2011) (see also Kahneman, 2002). For cognitive (mechanistic) researchers, this involves defining outcomes or effects functionally (i.e., as the causal impact of the environment on behavior). Doing so circumvents many of the "behavior-as-proxy" challenges highlighted previously, illuminates the environmental causes of behavior and facilitates the discovery of new mental mediators. For functionally orientated researchers, the mechanistic approach may provide access to novel methodologies and previously unconsidered domains that can be used to further refine existing functional theories and models. Furthermore, building a rich body of functional knowledge may be profitable for both traditions given that (a) it would enforce much needed constraints on mental explanations and help facilitate the selection of mental accounts from one another, (b) provide information about environment–behavior interactions that could be used to refine existing functional theories as well as (c) lead to new conceptual tools for describing in non-mental terms behavioral effects and the conditions under which these effects occur. We hope that the analysis and discussion offerred throughout the current paper not only clarify the nature of and relation between different approaches to implicit cognition but also stimulate continued dialog and progression within and between these respective traditions.

Acknowledgments

This paper was funded by a scholarship from Irish Research Council for Science, Engineering and Technology (IRCSET) to the first author. We would like to thank Jan De Houwer for his invaluable insight and feedback on many of the topics discussed in this paper.

References

Albarracin, D., Glasman, L. R., & Wallace, H. M. (2004). Survival and change in judgments: A model of activation and comparison. In: M. P. Zanna (Ed.), *Advances in experimentalsocial psychology*, *36* (pp. 251–315). San Diego: Academic Press.

Agerström, J., & Rooth, D. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*, *96*, 790–805.

Ames, S. L., Grenard, J. L., Thush, C., Sussman, S., Wiers, R. W., & Stacy, A. W. (2007). Comparison of indirect assessments of association as predictors of marijuana use among at-risk adolescents. *Experimental and Clinical Psychopharmacology*, *15*(*2*), 204–218.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*, 143–148.

Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In: S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 348–388). New York: John Wiley & Sons.

Bar-Anan, Y. & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin*. Advance on-line publication. http://dx.doi.org/10.1177/0146167212446835.

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, R. S., Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed.). Hillsdale, NJ: Erlbaum.

Barnes, D., Hegarty, N., & Smeets, P. (1997). Relating equivalence relations to equivalence relations: A relational framing model of complex human functioning. *The Analysis of Verbal Behavior*, *14*, 1–27.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. *Psychological Record*, *50*(*2*), 251–265.

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's *Verbal Behavior:* A possible synthesis. *The Behavior Analyst*, *23*, 69–84.

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

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., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record*, *58*, 497–515.

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., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes towards meat and vegetables in vegetarians and meat-eaters. *The Psychological Record*, *60*, 287–306.

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

Bentall, R. P., Dickins, D.W, & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology*, *46*, 187–1214.

Biglan, A., & Hayes, S. C. (1996). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology: Current Scientific Perspectives*, *5*, 47–57.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(*4*), 1061–1071.

Brown, A. S., Gray, N. S., & Snowden, R. J. (2009). Implicit measurement of sexual preferences in child sex abusers: Role of victim type and denial. *Sexual Abuse: A Journal of Research and Treatment*, *21*, 166–180.

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.

Carpenter, K. M., Martinez, D., Vadhan, N. L., Barnes-Holmes, D., & Nunes, E. V. (2012). Performance based measures of attentional bias and relational responding are associated with behavioural treatment outcome for cocaine dependence. *The American Journal of Drug and Alcohol Abuse*, *38*(*2*), 146–154.

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.

Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science*. Boston: Authors Cooperative.

Choma, B. L., & Hafer, C. L. (2009). Understanding the relation between explicitly and implicitly measured political orientation: The moderating role of political sophistication. *Personality and Individual Differences*, *47*, 964–967.

Cserjesi, R., Vermeulen, N., Luminet, O., Marechal, C., Nef, F., Simon, Y., et al. (2010). Explicit vs. implicit body image evaluation in restrictive anorexia nervosa. *Psychiatry Research*, *175*, 148–153.

Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The Psychological Record*, *59*, 591–620.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The Iterative Reprocessing Model: A multi-level framework for attitudes and evaluation. *Social Cognition*, *25*, 736–760.

Czopp, A. M., Monteith, M. J., Zimmerman, R. S., & Lynam, D. R. (2004). Implicit attitudes as potential protection from risky sex: Predicting condom use with the IAT. *Basic and Applied Social Psychology*, *26*, 227–236.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

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. (2008). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In: R. Petty, R. H. Fazio, & P. Brionol (Eds.), *Attitudes: Insights from the new implicit measures*. Erlbaum.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*, 443–451.

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. (2011). 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., & Moors, A. How to define and examine implicit processes? In R. Proctor & J. Capaldi (Eds.). *Implicit and explicit processes in the psychology of science*. Oxford University Press, in press.

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., & Moors, A. (2010). Implicit measures: Similarities and differences. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

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.

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and fragile: A new look at the automaticity of negation processing. *Experimental Psychology*, *56*, 434–446.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *Psychological Record*, *56*, 83–103.

Dotsch, R., & Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, *44*, 1194–1198.

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.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540.

Dymond, S., & Barnes, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. *The Psychological Record*, *46*(*2*), 271–300.

Dymond, S., & Barnes, D. (1997). Behavior analytic approaches to self-awareness. *The Psychological Record*, *47*, 181–200.

Dymond, S. & Roche, B. (Eds.). *Advances in Relational Frame Theory & Contextual Behavioral Science: Research & Application*. Oakland, CA: New Harbinger Publications, in press.

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

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637.

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.

Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: The effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology*, *87*, 557–572.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Test. *Basic and Applied Social Psychology*, *27*, 307–316.

Fields, L., & Moss, P. (2007). Stimulus relatedness in equivalence classes: Interaction of nodality and contingency. *European Journal of Behavior Analysis*, *8*, 141–159.

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

Friese, M., Wänke, M., & Plessner, H. (2006). Implicit consumer preferences and their influence on product choice. *Psychology & Marketing*, *23*, 727–740.

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

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

Garcia-Marques, L., & Ferreira, M. (2011). Friends and foes of theory construction in psychological science: Vague dichotomies, unified theories of cognition, and the new experimentalism. *Perspectives on Psychological Science*, *6*(*2*), 192–201.

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

Gaudiano, 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., & Creighton, L. A. Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition*. New York: Oxford University Press, in press.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion*, *24*, 1008–1025.

Gawronski, B., & De Houwer, J. Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed.). New York: Cambridge University Press, in press.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370–377.

Gawronski, B., LeBel, E. P., Peters, K. R., & Banse, R. (2009). Methodological issues in the validation of implicit measures. Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin*, *135*, 369–372.

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., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216–240). New York: Guilford Press.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*, 535–542.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford 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.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.

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, 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. P., Seibt, B., & Banaji, M. H. (2006). Easier done than undone: Asymmetries in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.

Gross, A. C., & Fox, E. J. (2009). Relational frame theory: An overview of the controversy. *The Analysis of Verbal Behavior*, *25*, 87–98.

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

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, *11*, 2351–2355.

Hayes, S. C. (1993). Analytic goals and the varieties of scientific contextualism. In: S. C. Hayes, L. J. Hayes, H. W. Reese, & T. R. Sarbin (Eds.), *Varieties of scientific contextualism* (pp. 11–27). Reno, NV: Context Press.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavior therapy. *Behavior Therapy*, *35*, 639–665.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). New York: Plenum Press.

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

Hayes, S. C., & Bissett, R. (1998). Derived stimulus relations produce mediated and episodic priming. *The Psychological Record*, *48*, 617–630.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior–behavior relations, and abehavior–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.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In: S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A postSkinnerian account of human language and cognition* (pp. 21–49). New York: Plenum.

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*(*2*), 207–227.

Hermsen, B., Holland, R. W., & van Knippenberg, A. (2006). The happy act on impulse, the sad think twice: Mood as a moderator of the impact of implicit and explicit attitudes on behavior. *Unpublished manuscript*.

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

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, *30*, 161–172.

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.

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. A Functional Approach to the study of implicit cognition: The Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. In Dymond, S & Roche, B. (Eds.). *Advances in Relational Frame Theory & Contextual Behavioral Science: Research &Application*. Oakland, CA: New Harbinger Publications, in press.

Hussey, I., & Barnes-Holmes, D. The IRAP as a measure of implicit depression and the role experiential avoidance. *Cognitive and Behavioral Practice*, in press.

Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the time course of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*, 1267–1280.

Ju, W. C., & Hayes, S. C. (2008). Verbal establishing stimuli: Testing the motivative effect of stimuli in a derived relation with consequences. *The Psychological Record*, *58*, 339–363.

Kahneman, D. (2002). Autobiography. Retrieved October 5, 2006, from the Nobel Foundation Web site: 〈http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-autobio.html〉.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871–888.

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes*, *24*, 219–245.

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility order effects: New evidence for a task-set switching account of the implicit association test effect. *Personality and Social Psychology Bulletin*, *31*, 208–217.

Lee, S., Rogge, R. D., & Reis, H. T. (2010). Assessing the seeds of relationship decay: Using implicit evaluations to detect the early stages of disillusionment. *Psychological Science*, *21*(*6*), 857–864.

Leslie, J. C., Tierney, K. J., Robinson, C. P., Keenan, M., Watt, A., & Barnes, D. (1993). Differences between clinically anxious and nonanxious participants in a stimulus equivalence training task involving threat words. *The Psychological Record*, *43*, 153–161.

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

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, *34*, 199–249.

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.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*, 405–415.

Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. J. (2001). Heart versus reason in condom use: Implicit vs. explicit attitudinal predictors of sexual behavior. *Zeitschrift fur Experimentelle Psychologie*, *48*, 161–175.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*, 435–442.

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Social group association cues: Forming implicit and explicit attitudes toward individuals. *Journal of Personality and Social Psychology*, *94*, 792–807.

McGlinchey, A., Keenan, M., & Dillenburger, K. (2000). Outline for the development of a screening procedure for children who have been sexually abused. *Research on Social Work Practice*, *10*, 722–747.

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

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2007). Deictic relational complexity and the development of deception. *The Psychological Record*, *57*(*4*), 517–531.

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

Merwin, R. M., & Wilson, K. G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *Psychological Record*, *55*, 561–575.

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

Mogg, K., Bradley, B. P., Field, M., & De Houwer, J. (2003). Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction*, *98*, 825–836.

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* (pp. 19–37). NY: Guilford Press.

Nicholson, E., & Barnes-Holmes, D. (2012). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust-propensity and -sensitivity in OCD. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(*3*), 922–930.

Nicholson, E., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record*, *62*, 263–278.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., et al. (2006). Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, *39*, 299–321.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Nock, M. K., Park, J. L., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the "suicidal mind": Implicit cognition predicts suicidal behavior. *Psychological Science*, *21*, 511–517.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science*, *16*, 65–69.

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.), *Automatic processes in social thinking and behavior* (pp. 265–292). Psychology Press.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*, 553–594.

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'Reilly, A., Roche, B., Ruiz, M., Tyndall, I., & Gavin, A. (2012). The Function Acquisition Speed Test (FAST): A behavior-analytic implicit test for assessing stimulus relations. *The Psychological Record*, *62*, 507–528.

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.

O'Toole, C., & Barnes-Holmes, D. (2009). Three chronometric indices of relational responding as predictors of performance on a brief intelligence test: The importance of relational flexibility. *The Psychological Record*, *59*, 119–132.

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

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In: R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19–63). New York: Psychology Press.

Paladino, M., & Castelli, L. (2008). On the immediate consequences of intergroup categorization: Activation of approach and avoidance motor behavior toward ingroup and outgroup members. *Personality and Social Psychology Bulletin*, *34*(6), 755–768.

Parling, T., Cernvall, M., Stewart, I., Barnes-Holmes, D., & Ghaderi, A. (2012). Using the Implicit Relational Assessment Procedure to compare implicit pro-thin/anti-fat attitudes of patients with anorexia nervosa and non-clinical controls. *Eating Disorders: The Journal of Treatment and Prevention*, *20*(*2*), 127–143.

Paulhus, D. L. (1989). Socially desirable responding. In: D. M. Buss, & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 201–209). New York: Springer-Verlag.

Payne, K., Burkley, M., & Stokes, M. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*, 16–31.

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.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going?. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1–15). New York: Guilford Press.

Pepper, S. C. (1942). *World hypotheses: A study in evidence*. Berkeley, CA: University of California Press 1970.

Perez, E. O. (2010). Explicit evidence on the import of implicit attitudes: The IAT and immigration policy judgments. *Political Behavior*, *32*, 517–545.

Perkins, A., & Forehand, M. (2010). Implicit social cognition and indirect measures in consumer behavior. In: B. Gawronski, & K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In: B. Gawronski, & K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford pp. 255–277.

Peters, K. R. (2011). *Carving cognition at its joints: Insights from the interaction between explicit and implicit social cognition*. Unpublished Doctoral Dissertation, University of Western Ontario, Canada.

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.

Petty, R. E., Briñol, P., & DeMarre, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, *25*, 657–686.

Polaschek, D. L. L., Bell, R. K., Calvert, S. W., & Takarangi, M. K. T. (2010). Cognitive-behavioural rehabilitation of high-risk violent offenders: Investigating treatment change with explicit and implicit measures of cognition. *Applied Cognitive Psychology*, *24*(*3*), 437–449.

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.

Quiñones, J. (2008). *Relational coherence and transformation of function in ambiguous and unambiguous relational networks*. Unpublished Doctoral Dissertation, University of Nevada, Reno.

Ranganath, 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.

Reilly, T., Whelan, R., & Barnes-Holmes, D. (2005). The effect of training structure on the latency responses to a five-term linear chain. *The Psychological Record*, *55*(*2*), 233–249.

Richeson, J.A, & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–423.

Richetin, J., Perugini, M., Prestwich, A., & O'Gorman, R. (2007). The IAT as a predictor of spontaneous food choice: The case of fruits versus snacks. *International Journal of Psychology*, *42*, 166–173.

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes-Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record*, *50*, 267–292.

Roche, B. T., Kanter, J. W., Brown, K. R., Dymond, S., & Fogarty, C. C. (2008). A comparison of "direct" versus "derived" extinction of avoidance responding. *The Psychological Record*, *58*, 443–464.

Roche, B., Linehan, C., Ward, T., Dymond, S., & Rehfeldt, R. (2004). The unfolding of the relational operant: A real-time analysis using electroencephalography and reaction time measures. *International Journal of Psychology & Psychological Therapy*, *4*(*3*), 587–603.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2010). Anti-fat, pro-slim, or both? Using two reaction time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology*, *15*, 416–425.

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.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, *114*, 363–375.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Eliminating the influence of recoding in the Implicit Association Test: The Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology*, *62*(*1*), 84–98.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139–165.

Rudman, L. A., Phelan, J. E., & Heppen, J. B. (2007). Developmental sources of implicit attitudes. *Personality and Social Psychology Bulletin*, *33*, 1700–1713.

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.

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior*, *72*, 117–137.

Scarabis, M., Florack, A., & Gosejohann, S. (2006). When consumers follow their feelings: The impact of affective or cognitive focus on the basis of consumers' choice. *Psychology & Marketing*, *23*, 1005–1036.

Schnabel, K., Banse, R., & Asendorpf, J. B. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology*, *53*, 69–76.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*(*5*), 638–656.

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In: A. Tesser, & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes*, 1 (pp. 436–457). Oxford, UK: Blackwell.

Shook, N. J., & Fazio, R. H. (2008). Roommate relationships: A comparison of interracial and same-race living situations. *Group Processes and Intergroup Relations*, *11*, 425–437.

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

Sinclair, S., Dunn, L., & Lowery, B. (2005). The relationship between parental racial attitudes and children's implicit prejudice. *Journal of Experimental Social Psychology*, *41*, 283–289.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, *52*, 270–277.

Smith, C. T., De Houwer, J. & Nosek, B. A. Consider the source: Persuasion of implicit evaluations is moderated by source credibility. Personality and Social Psychology Bulletin, in press.

Smith, E. R., & DeCoster, 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.

Snowden, R. J., Craig, R. L., & Gray, N. S. (2011). Indirect behavioral measures of cognition among sexual offenders. *Journal of Sex Research*, *48*(*2–3*), 192–217.

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. (2004). Relational frame theory and analogical reasoning: Empirical investigations. *The International Journal of Psychology and Psychological Therapy*, *4*(*2*), 241–262.

Stewart, I., Kelly, G., McHugh, L., Barnes-Holmes, D., & O'Hora, D. Second order contextual control over non-arbitrary relational responding: An empirical investigation and preliminary model of pragmatic verbal analysis, submitted for publication.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Teachman, B. A. (2007). Evaluating implicit spider fear associations using the go/no-go association task. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 156–167.

Teachman, B., & Brownell, K. (2001). Implicit associations toward obese people among treatment specialists: Is anyone immune? *International Journal of Obesity*, *25*, 1–7.

Teachman, B. A., Cody, M. W., & Clerkin, E. M. (2010). Clinical applications of implicit measures. In: K. Payne, & B. Gawronski (Eds.), *Handbook of implicit social cognition: Measurement, Theory, and Applications* (pp. 489–521). New York: Guilford Press.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: The Single Block IAT. *European Journal of Psychological Assessment*, *24*, 237–245.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). Practical guide to Implicit Association Task and related tasks. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York: Guilford Press.

Timmins, L., Barnes-Holmes, D., & Cullen, C. (2012). Measuring Implicit Sexual Response Biases to Nude Male and Female Pictures in Gay and Heterosexual Men. *Unpublished manuscript*.

Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior*, *85*, 349–369.

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

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally orientated self-control. *Perspectives of Psychological Science*, *5*, 243–262.

Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, *59*, 371–388.

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

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

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

Wang, T., McHugh, L., & Whelan, R. (2012). A test of the discrimination account in equivalence class formation. *Learning and Motivation*, *43*, 8–13.

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record*, *41*(1), 33–50.

Watson, J. B. (1924). *Behaviorism*. New York: People's Institute Publishing Company.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York: Guilford Press.

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.

Wiers, R. W., Houben, K., Roefs, A., de Jong, P., Hofmann, W., & Stacy, A. W. (2010). Implicit cognition in health psychology: Why common sense goes out of the window. In: B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 463–488). New York: The Guilford Press.

Wilson, T. D. (2009). Know thyself. *Perspectives on Psychological Science*, *4*, 384–389.

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.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality & Social Psychology*, *81*, 815–827.

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition and Emotion*, *22*, 1187_1197.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, *50*, 125–144.