University of Mississippi

eGrove

Electronic Theses and Dissertations   Graduate School

2014

Stop Making Sense: Exploring Basic Properties and Clinical Applications of Coherence

Michael James Bordieri

Follow this and additional works at: https://egrove.olemiss.edu/etd

Part of the Clinical Psychology Commons

Recommended Citation

STOP MAKING SENSE: EXPLORING BASIC PROPERTIES AND CLINICAL

APPLICATIONS OF COHERENCE

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
The University of Mississippi

by

MICHAEL J. BORDIERI

April 2013

ABSTRACT

This study explored the ways in which people make sense of ambiguous tasks and the degree to which people prefer contexts where coherent responding is possible. Relational frame theory contains a foundational assumption that coherence (i.e., making sense) is reinforcing for verbally competent humans. That is, it is assumed that humans relate ambiguous stimuli in ways that go together because they have an extensive learning history where others have given praise, positive attention, and other reinforcement for this behavior. This study was designed to empirically investigate this core assumption of relational frame theory by analyzing response patterns to ambiguous stimuli and by assessing whether participants displayed a preference towards coherent contexts. Obtained findings revealed that the majority of participants responded to ambiguous stimuli in ways that were internally consistent and coherent in the absence of any programmed contingencies. Many participants also displayed a preference toward contexts where coherent responding was possible and a small subset of participants persisted in this preference even when it was increasingly costly to do so. Reports of frustration obtained throughout the preparation were moderated both by performance in study tasks and by measures of cognitive fusion and psychological inflexibility. The major theoretical contributions of these findings as well as applied implications were discussed.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of the Department of Psychology and the Mississippi Center for Contextual Psychology workgroup, which provided an atmosphere of intellectual stimulation, personal growth, and practical assistance throughout my time at the University of Mississippi. I am particularly appreciative of my advisor, Dr. Kelly Wilson, and Dr. Kate Kellum, who both encouraged me to grow as a researcher, clinician, teacher, and as a person. I would also like to thank my other committee members, Dr. John Young and Dr. Mark Van Boening, for their thoughtful comments that helped refine this project. Finally, I would like to thank Kerry Whiteman, who not only served as a reliability coder and helped proofread this manuscript, but also provided unending support, kindness, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Tiger got to hunt,
Bird got to fly;
Man got to sit and wonder, "Why, why, why?"

Tiger got to sleep,
Bird got to land;
Man got to tell himself he understand. (Vonnegut, 1963, p. 88)

The ability to make sense of the world by relating thoughts and ideas together is a defining feature of human behavior. As Vonnegut astutely notes, the acts of asking "why" and generating understanding appear to be ubiquitous features of human behavior. Sense making also appears central to the scientific endeavor. Prominent scientific historian Peter Dear contends that science is most commonly viewed as, " a natural philosophy, which strives to give an account of nature – to make sense of it" (2006, p. 2). This received view of science, as it is referred to by Wilson and colleagues (in press), subscribes to a realist philosophy of science that seeks to develop scientific theories that explain the true nature of the world (cf. Popper, 2002; Putnam, 1975). Alternative philosophies of science such as James' (1907) pragmatism, Skinner's (1974) radical behaviorism, and functional contextualism (Hayes, 1993; see also Biglan & Hayes, 1996), reject the ontological claims made by the realist. These approaches embrace effective action as their truth criterion and view the scientific endeavor not as an exercise in uncovering truths, but as an attempt on the part of scientists to interact more effectively with the world (Pepper, 1942). Despite possessing fundamentally different truth criterions, both scientific philosophies hold sense making as the primary activity of scientists.

Realists make sense of the world by seeking to understand its true nature while contextualists make sense of the world by seeking ways to interact more effectively with it.

Studying sense making from a scientific perspective therefore presents a conundrum. How can a scientist make sense of sense making behavior when the tools at their disposal are identical to the target of inquiry? This problem of subjectivity of observation and access to mental events has been wrestled with by psychology since its inception as a discipline (cf. Boring, 1953). Proponents of the received view of science address this problem by operationally defining a set of agreed upon theoretical terms that are assumed to genuinely refer to concepts in the world (see Wilson, 2001 and Wilson, Whiteman, & Bordieri, in press for a more in-depth discussion of operationalism). In the case of mental events such as sense making, operationalization is accomplished by appealing to the native capacities of the human mind. For instance, Chomsky's linguistic nativism (1965) remains a foundational assumption of contemporary cognitive theory to this day (Samuels, 2004). Within this perspective, sense making is assumed to be an innate feature of human minds that is amenable to scientific study in an effort to determine the underlying sense making structures and mechanisms. Recent theoretical and empirical efforts have been launched from contemporary cognitive theory that study sense making in domains such as information management (Dervin, 1998), military command structures (Jensen, 2009), human-computer interactions (Pirolli & Russell, 2011), and organizational science (Weick, Sutcliffe, & Obstfeld, 2005).

Radical behaviorism and the contextualist perspective offer another approach to this problem of subjectivity and access. In particular, Skinner (1945) provides an alternative approach to the operationalization of scientific terms such as sense making. Instead of looking for the meaning of sense making as a thing whose properties need to be uncovered and

articulated, Skinner argues that the focus should be placed on identifying and analyzing the conditions under which we use the term sense making. Put succinctly, "meaning, contents, and references are to be found among the determiner, not among the properties, of response (1945, p. 271). To Skinner, the meaning of sense making is not to be found in an agreed upon definition, rather it is to be found by exploring the variety of contexts under which scientists use the term. What follows is an analysis exploring the scientific contexts where the term sense making is used.

The study of human sense making behavior is a broad domain of study and the phenomenon has been investigated from a variety of perspectives (Wray, 2011). Various terminologies have been used to describe it such as forming a self narrative (Pennybaker & Seagal, 1999; Roe & Davidson, 2005), making meaning (Janoff-Bulman & Frieze, 1983), meaning making (Lips-Wiersma, 2002) post traumatic growth (Park, Riley, & Snyder, 2012), developing a sense of coherence (Drageset, Espehaug, & Kirkevold, 2012; Kazmierczak, Strelau, & Zawadzki, 2012) reason giving (Hayes, Barnes-Holems, & Roche, 2001), story telling (Wilson & DuFrene, 2008), and rumination (Papageorgiou & Wells, 2003).

Consistent across all these uses of the term is that sense making is something that a person does. That is, sense making can be conceptualized as a behavior, and thus amenable to a behavioral analysis. The behavioral tradition has a rich history of providing useful analyses of complex operant behaviors such as variability (Page & Neuringer, 1985) and creativity (Winston & Baker, 1985). Therefore, it appears appropriate to address the complex behavior of sense making from a behavioral perspective.

Making Sense of Sense Making

Conceptualized broadly, all behavior of organisms can be considered sense making. A foundational assumption of behavior analysis is that behavior is orderly controlled by the current

context (including antecedents and consequences) and the organism's learning history. Thus, Thorndike's cats learning how to escape the puzzle box and Skinner's rats learning to press the lever for food are clear examples of animals "making sense" of the world (Skinner, 1938; Thorndike, 1898 as cited in Chance, 1999). However, this type of non-verbal sense making is of a different quality than the sense making displayed by verbally competent humans. As Hayes (1997) notes, human sense making and knowing is a verbal process that has as a defining feature a sense of perspective and self-awareness. That is, while one of Throndike's cats may be able to escape from a puzzle box, it is unable to "know" that it did so. Skinner explains the distinction thusly:

> There is a...difference between behaving and reporting that one is behaving or reporting the causes of one's behavior. In arranging conditions under which a person describes the public or private world in which he lives, a community generates that very special form of behavior called knowing. . . Self-knowledge is of social origin. (1974, p. 34-35)

This development of self-awareness and the ability to "know" and "make sense" of the world is purported to be achieved by language (Hayes, 1997).

Sense making as verbal behavior. Language has long been an important area of interest within the behavioral tradition with Kanter (1926; 1929) and Skinner (1957) both providing detailed theoretical accounts of language and verbal behavior. These early theoretical accounts, and in particular Skinner's *Verbal Behavior* (1957), have led to a modest generation of empirical work (Dixon, Small, & Rosales, 2007; Dymond, O'Hora, Whelan, & O'Donovan, 2006). In contrast, contemporary behavioral accounts of language such as stimulus equivalence (Sidman, 1971; 1994), the naming hypothesis (Horne & Lowe, 1996), and relational frame theory (Hayes, Barnes-Holmes, & Roche, 2001) have generated a substantial body of empirical evidence over the past forty years. In particular, relational frame theory has generated over sixty published

empirical tests of its core tenets, making it an ideal framework for conducting a cotemporary behavioral analysis of sense making and coherence (Dymond, May, Munnelly, & Hoon, 2010; Hayes, Luoma, Bond, Masuda, & Lillis, 2006).

Relational Frame Theory (RFT), a theoretical account of language and cognition, conceptualizes verbal behavior as the ability to relate arbitrary concepts together and respond according to symbolic relations (Hayes, Barnes-Holmes, & Roche, 2001). It holds at its core concept that language and cognition are behavioral events (i.e., generalized operants) that are comprised of arbitrarily applicable relational frames (Hayes, et al., 2001). These relational frames are arbitrary in the sense that they are not based on formal properties of stimuli. Rather, they are based on functional relationships between stimuli (i.e., relational frames). Derived relational responding (DRR) is a three-term contingency where an individual who has a history of differential reinforcement correlated with a contextual cue emits a relational response based on the presence of said cue (Healy, Barnes-Holmes, & Smeets, 2000).

As an example, consider an individual who has a history of being reinforced for matching the word lemon to the word jibjar and the word lemon to an actual lemon. RFT accounts for three behaviors that emerge from these two trained equivalence relationships. Firstly, an individual will match the word jibjar to the word lemon and match an actual lemon to the word lemon. This process is referred to as mutual entailment. Secondly, an individual will match the word jibjar to an actual lemon and vice versa (i.e., an actual lemon to jibjar). This process is referred to as combinatorial entailment. See Figure 1 for a graphical depiction of mutual and combinatorial entailment. Finally, a transformation of stimulus functions occurs such that functions of a stimulus within the relational network may transfer to other stimuli. As an illustration, imagine biting into a ripe, juicy, jibjar. It is quite possible that some stimulus

5

functions of an actual lemon (e.g. sour, tart, etc.) may have transferred to the word "jibjar" as you read this paragraph occasioning behavioral responses such as salivating or puckering. These three elements (mutual entailment, combinatorial entailment, and transformation of stimulus functions) form the three basic components of RFT (Hayes et al., 2001).

```
                        ┌──────────┐
                        │          │
                        │  lemon   │      ───────  directly trained
                        │          │      - - - -  mutual entailment
                        └──────────┘      ········  combinatorial entailment
              ┌──────────┐      ┌──────────┐
              │          │      │          │
              │  jibjar  │<····>│  (lemon) │
              │          │      │          │
              └──────────┘      └──────────┘
```
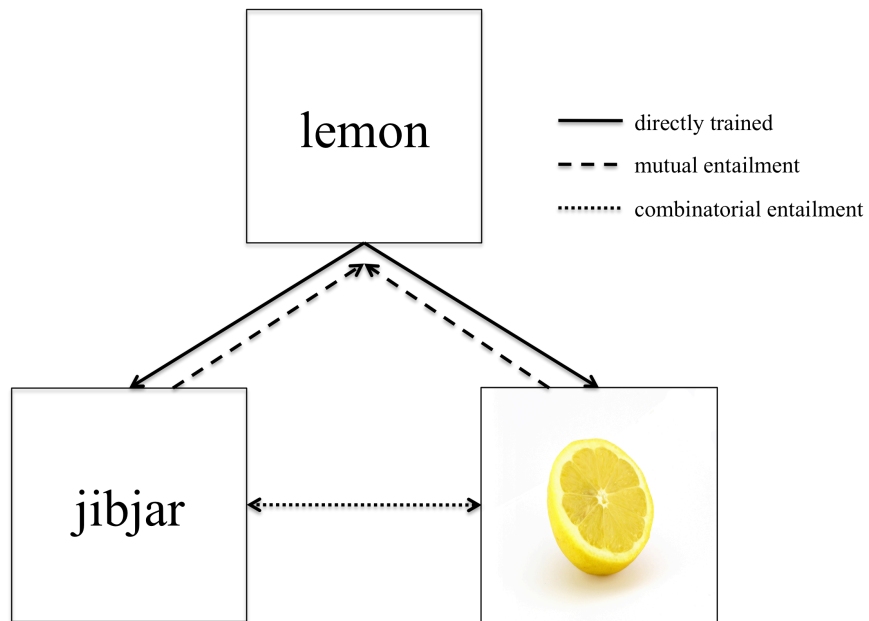
Figure 1. Graphical depiction of mutual and combinatorial entailment.

A growing body of empirical support has emerged to support these basic components of RFT. Basic research studies have demonstrated the emergence of mutual and combinatorial entailment (Hayes, Thompson, & Hayes, 1989; Steel & Hayes, 1991). In addition, transformation of stimulus functions has been demonstrated in a variety of different experimental preparations (Barnes-Homles & Keenan, 1993; Dougher, Auguston, Markham, Greenway, & Wulfret, 1994; Dougher, Hamilton, Fink, & Harrington, 2007; Hayes, Kohlenberg, & Hayes, 1991; Wulfert & Hayes, 1988).

Coherence as a reinforcer: An untested assumption of verbal behavior. Despite strong evidence of support for its core tenets, one of the basic assumptions of RFT remains largely untested. Namely, the assumption that verbally competent humans engage in coherent

derived relational responding because doing so has been reinforced and is, therefore, reinforcing. This assumption is embedded into current theoretical accounts of RFT:

> Coherence and utility are enough to maintain verbal relations once they are established. Detecting that one is deriving coherent and explainable relational networks (e.g., learning that one is "right" or "making sense") or that relating events is leading to effective outcomes (e.g., learning that one has "solved the problem") and similar processes provide continuous reinforcement for the process of relational framing. (Hayes, Strosahl, & Wilson, 2011, pp. 51-52)

The entire RFT analysis rests on this assumption, as the theory is incomplete without an adequate explanation for the conditions under which verbal behavior (i.e., derived relational responding) is trained and maintained. The major proponents of RFT acknowledge this limitation, stating that empirical identification of the precise learning histories involved in the acquisition and maintenance of derived relational responding (DRR) remains to be fully explicated (Hayes, Fox, Gifford, & Wilson, 2001, p. 28). What follows is a brief review of the available evidence for the environmental conditions linked to acquisition and maintenance of DRR.

There is growing evidence supporting the RFT account of how derived relational responding is acquired. For example, studies have shown that derived relational responses are absent in non-verbal children (Barnes, McCullagh, & Kennan, 1990; Devany, Hayes, & Nelson, 1986). In addition, the acquisition of the DRR repertoire is largely consistent with the developmental trajectory of language (Lipkens, Hayes, & Hayes, 1993) and the repertoire itself can be directly trained (Luciano, Gomez-Becerra, & Rodriquez-Valverde, 2007). Furthermore, more recent studies have demonstrated that providing training in core DRR repertoires increases complex skills such as intelligence (Cassidy, Roche, & Hayes, 2011) and perspective taking (Rehfeldt, Dillen, Ziomek, & Kowalchuk, 2007; Weil, Hayes, & Capurro, 2011). Likewise, deficits in DRR repertoires have been associated with lower levels of performance on a standardized measure of intelligence (O'Hora, Pelaez, & Barnes-Holmes, 2005).

While evidence for the acquisition of derived relational responding is increasing, there remains only limited evidence for consequences that maintain DRR. RFT theorists posit that the behavior of deriving coherent relational networks (i.e., relations that make sense) may be self-reinforcing (cf. Hayes et al., 2011, p. 52). However, direct evidence for the reinforcing functions of coherence (i.e., sense-making) remains largely anecdotal or interpretative. For example, in describing an early form of a testing procedure, Barnes-Holmes and colleagues (2001) observed that most subjects withdrew from the study because they couldn't figure out how to respond. When they changed the study to allow for clear responding, participants reacted positively and persisted in the task. Based on this observation, they concluded that, "coherence or sense-making appears to function as a powerful reinforcer for relational activity" (Barnes-Holmes, Hayes, Dymond, & O'Hora, 2001, p. 70). This observation is not unique to contemporary behavioral thought. For example, Skinner hypothesized that the effects of thinking are reinforcing (1953, pp. 242-256) and provided a theoretical account for how responses initially reinforced by the social verbal community can recede into the private domain (Skinner, 1945). Despite these assertions, direct empirical testing of the reinforcing properties of coherence remains largely absent from the literature.

Empirical Investigations of Coherence

Experimental evidence suggestive of the reinforcing properties coherence. While direct evidence is limited, indirect evidence of the reinforcing properties of coherence is abundant. Much of the general empirical support for RFT is suggestive of coherence serving as a reinforcer. In fact, the primary means of empirically demonstrating DRR is during testing tasks in which no programmed reinforcers are provided. Given the absence of other contingencies such as social praise or monetary incentives, it stands to reason that coherence (i.e., making

sense of the task) may be the contingency controlling participant responding. As an example, consider multiple empirical demonstrations where participants displayed a clear preference towards a slot machine whose color was related relationally to the function of "greater than" compared to a slot machine whose color was related relationally to the function of "less then" (Hoon, Dymond, Jakcson & Dixon, 2008; Nastally, Dixon, & Jackson, 2010; Zlomke & Dixon, 2006). In all three of these preparations, the direct payout odds of the slot machines were equal, suggesting that the consequence controlling participants' choice behavior may be coherence (i.e., making sense of the task by picking the machine associated with bigger).

Wilson and Hayes (1996) provide additional evidence suggestive of the role coherence plays in derived relational responding. In their preparation designed to explore resurgence of derived relations, they initially trained stable stimulus classes (early training) and then trained new relations among the stimulus classes that contradicted the early training (late training). They then directly punished responses consistent with the late training and assessed how participants reacted. Sixteen of the 23 total subjects displayed response patterns consistent with their early training providing direct evidence of the resurgence of the previously derived relations. These findings can also be interpreted as evidence of coherence as the majority of participants adopted a previously reinforced response strategy to "make sense" of the ambiguous task. Of particular note is the fact that 22 out of the 23 participants were identified as displaying a clear, internally consistent response patterns when faced with the ambiguous task. That is, once participants decided on how to respond, they persisted in the same pattern with little deviation. This suggests that responding in a consistent and coherent way may function as a reinforcer. A recent replication and extension of this preparation by Doughty, Kastner, & Bismark (2011) found similar results. In their discussion, the authors directly speculated as to

the role of coherence as a reinforcer stating, "perhaps reinforcing consequences were produced automatically by behaving in a manner consistent with the baseline discriminations necessary for stimulus equivalence" (2011, p. 154).

Evidence from literature investigating the stability of derived stimulus relations also lends indirect support to coherence functioning as a reinforcer. Of primary interest is the finding that derived stimulus relations persist over time. For example, Saunders, Wachter, and Spradlin (1988) demonstrated that derived stimulus relations remained intact for five months without any review or re-training. The susceptibility of derived stimulus relations to interruption provides additional support. Some studies suggest that interruption has little effect on derived relations (Saunders, Saunders, Kirby & Spradlin, 1988; Spradlin, Saunders, & Saunders, 1992) while others studies suggest clear evidence of partial class disruption (Pilgrim & Galizio, 1995; Pilgrim, Chambers, & Galizio, 1995). Of particular interest to the current investigation are the findings of Pilgrim, Chambers, and Galizio (1995) who observed that young children are more susceptible to class disruption than adults. In discussing their findings the authors suggest that, "'consistency' in responding is a less well-established property, or higher order class, for young children"(1995, p. 253). That is, adults have a greater learning history of responding in a coherent manner and thus would be expected to be less susceptible to disruptions in their relational responding repertoire.

The area of rule-governed behavior is also relevant to an analysis of coherence. The term rule governed behavior was first used by Skinner (1966) as a means to describe behavior controlled by a contingency specifying stimulus. Put simply, this area of research is interested in the ways in which verbal rules such as "I shouldn't try too hard because I'll just fail" come to control behavior (see Torneke, Luciano, & Salas, 2008 for a review and Hayes, 1989 for an in-

depth analysis). Of particular interest to an analysis of sense making behavior is the ways in which rule governed contingencies can override direct contingences. For example, Shimoff, Catania, & Mathews (1981) conducted a study where participants were either directly instructed via a verbal rule or shaped through contact with direct contingencies to press a lever slowly to earn points.   When the lever pressing contingencies were changed such that high rates of responding resulted in more points, participants who were given the verbal rule were largely insensitive to the change while participants whose behavior was directly shaped largely changed their responding. These findings, which suggest that rule-governed behavior is less sensitive to change than contingency governed behavior, have been replicated across a variety of experimental preparations (Galizio, 1979; Hayes, Brownstein, Zettle, Rosenfard, & Korn, 1986; Shimoff, Matthew, & Catania, 1986). Recent empirical investigations have also demonstrated that rule following behavior can be directly and indirectly trained (Kellum, 2009; Tarbox, Zuckerman, Bishop, Olive, & O'Hora, 2011).

These findings have significant implications in regards to the reinforcing properties of coherence. On a fundamental level, the relative insensitivity to direct contingencies exhibited during rule following suggests that participants must be attending to another stream of contingencies. There are three hypothesized ways in which rule following is purported to be acquired and maintained: pliance, tracking and augmentals (see Hayes & Wilson, 1993 for a detailed discussion).   Augmentals, verbal rules that establish or alter consequences of behavior, appear particularly relevant to coherence. Wilson & DuFrene (2008) have recently contended that values, a clinical application of augmentals, "establish predominate reinforcers . . . that are intrinsic in engagement in the valued pattern itself" (p. 64). This conceptualization suggests that the reinforcer for values rule-following lies within the congruence between one's behavior and

one's verbal formulation about said behavior. Put another way, it is reinforcing when our behavior coheres with statements about how we want to behave.

A final source of indirect evidence of the reinforcing properties of coherence lies within literature looking at the spontaneous generation of explanations. Skinner's (1936) work with the verbal summator, a device that generates ambiguous speech patterns, and projective tests such as the Rorschach ink blot test (see Weiner, 1994 for a contemporary review) both provide evidence of how humans work to generate meaning during ambiguous situations (cf. Skinner, 1953, p. 274). That is, when exposed to ambiguous speech sounds or visual stimuli, humans will often spontaneously emit meaningful and coherent responses; thus, "making sense" of the ambiguity. However, the propensity of verbally competent humans to engage in explanatory responses does not appear to be confined to ambiguous situations. For example, Peterson, Bettes, and Seligman (1982; as cited in Peterson & Seligman, 1984) found that 100% of college students who were asked to describe two negative life events provided at least one spontaneous causal explanation without being prompted to do so (n = 66).

A more recent study looking at political and moral opinion surveys also provides evidence of spontaneous generation of causal explanations. Hall, Johansson, and Strandberg (2012) used a deception procedure to reverse participants' endorsement of support on topics such as government surveillance of the Internet and the defensibility of Israel's use of violence against Hamas. Sixty nine percent of participants did not detect at least one of the reversed endorsements when asked by researchers to explain their responses. Of particular relevance to this investigation was the finding that over half of participants provided a coherent argument for the opposite of at least one of their original positions when their answers were surreptitiously

changed. These findings are suggestive of human's tendency to provide coherent explanations even in situations where explanations go against strongly held beliefs.

Taken as a whole, indirect empirical evidence of the reinforcing properties of coherence is substantial. Evidence looking at basic properties of derived stimulus relations, resurgence of derived stimulus relations, stability of derived stimulus relations, rule-governed behavior, and spontaneous generations of causal explanations all suggest that coherence (i.e., sense making behavior) functions as a reinforcer in verbally competent humans. However, it is important to temper this conclusion, as these empirical lines are merely suggestive. Direct assessment of the reinforcing functions of coherence is still needed.

Relatively direct evidence of the reinforcing properties of coherence. There is emerging evidence of coherence functioning as a reinforcer that is more direct in nature. Wray, Dougher, and Bullard (2008) exposed participants to both a solvable and unsolvable laboratory task where experimenter praise (i.e., positive feedback) was provided non-contingent on performance. They found that the majority of participants self-reported a preference towards the solvable task when rates of feedback were equal. In addition, they found that almost half of participants self-reported a preference towards the solvable task even when the rate of feedback was greater during the unsolvable task. These findings suggest that coherent contexts, in which sense making is possible, are generally preferred by verbally competent humans. In addition, the findings indicate that coherence may have functioned as a more powerful reinforcer than the directly programmed contingencies (i.e., positive feedback) for some participants. One major limitation of this study, which Wray (2011, p. 9) acknowledges, is that the assessment of preference was based solely on self-report and not on actual observations of behavior.

Wray (2011) conducted a second experiment to assess preference towards coherent, incoherent, or neutral contexts using direct observations of behavior as the primary measure of preference. Across three separate studies, Wray (2011) assessed participant responding with the third study yielding the greatest degree of experimental control. In this study, 17 participants were systematically exposed to three computerized matching to sample preparations where they were shown a sample stimulus and asked to select among three comparison stimuli. During the solvable matching to sample task, participants were reinforced for forming accurate conditional discriminations. During the neutral task, participants engaged in simple identity matching without programmed reinforcement. Finally, during the unsolvable matching to sample task, participants were asked to make novel conditional discriminations with inconsistent reinforcement provided to prevent class formation. After exposure to all three matching to sample conditions, participants were exposed to 10 concurrent choice and nine limited choice trials were they were asked to select which conditions they wanted to gain access to.

Results indicated a strong display of preference away from the unsolvable condition with only two of the 17 participants choosing the unsolvable task most frequently. Participants showed a greater degree of preference towards the neutral condition (53% of participants) than the solvable condition (24% of participants). These findings are typical of results obtained from the two other studies conducted by Wray (2011) and as a whole suggest that verbally competent humans prefer solvable or neutral conditions to unsolvable conditions in which coherent responding is not possible. The results from this experiment provide initial evidence suggestive of coherence functioning as a reinforcer.

It is important to note that the display of preference towards coherent conditions in Wray (2011) does not provide direct evidence that coherence itself is a reinforcer. Preference

assessments have displayed tremendous utility in identifying appetitive consequences (i.e., reinforcers) within applied behavior analysis (Fisher et al., 1992; Pace, Ivancic, Edwards, Iwata, & Page, 1985; Roane, Vollmer, Ringdahl, & Marcus, 1998). However, these preference assessment procedures are often confirmed by a reinforcer assessment, a procedure that directly assesses for higher rates of responding to preferred stimuli relative to non-preferred stimuli identified by the preference assessment. Absent findings from a direct reinforcer assessment, a procedure not employed in the Wray's (2011) study, it is not possible to state with certainly that coherence is, in fact, a reinforcer. Evidence of a clear preference towards contexts where coherence is possible strongly suggests that coherence has a reinforcing function for most participants. However, it remains an indirect measure.

Several other important dimensions of coherence remain unexplored. In particular, whether or not coherent responding will be displayed and preferred absent any programmed reinforcement remains an empirical question. In addition, an empirical assessment of the relative strength of coherence as a reinforcer is needed to determine whether or not the degree to which one is willing to work for coherence varies among verbally competent humans.

Applied Implications of Coherence

Setting aside the question of whether or not coherence is in itself reinforcing, there is a large body of empirical evidence suggesting that coherence can be adaptive or harmful depending on the context. That is, the behavior of sense making has been associated with both positive and negative consequences. This section will examine evidence of both and will conclude by exploring of a theoretical conceptualization of flexible sense making.

Adaptive sense making. Sense making can be extraordinarily adaptive as it allows us to interact more effectively with the world. In fact, the scientific enterprise as a whole can been

15

seen as the crowning example of the positive consequences generated by engaging in sense making behavior. One need not look further than the tremendous increases in average global life expectancy over the past two centuries (28.5 years in 1800 compared to 65.2 years in 1990; Riley, 2005) to see the result of sense making in medicine, engineering, public health, and other related disciplines. Sense making appears central to human success. So much so, that recent advances in evolutionary theory have contended that human language and cognition evolved as an adaptive trait precisely because it allows for social cooperation and more effective interactions with the environment (see Wilson & Wilson, 2007 for a detailed account and Hayes, Stroshal, & Wilson, 2011, pp. 16-21 for a discussion of the psychological implications). Put simply, humankind has benefited considerably from the development of the ability to engage in sense making behavior.

Sense making also serves a belonging function in our society. Skinner (1957, 1974) as well as Gergen and Gergen (1988) both contend that the social verbal community dictates which accounts of the world are considered accurate and which are not. As Wray (2011) notes, engaging in sense making behavior is a means of establishing credibility in the community, and doing so opens steady steams of reinforcement. Consider the behavior of the author constructing this paragraph. If the author arranges his ideas and sentences in ways that the reader deems "sensible" or "insightful," he is likely to be met with positive immediate consequences (e.g., the feedback "good job" or "that made sense"). In addition, continued sense making behavior will likely result in positive distal consequences (e.g., earning a Ph.D., career success, etc.). If the author instead arranges his ideas and sentences in ways that the reader deems "non-sensible" or "incoherent," he is likely to be met with aversive immediate consequences (e.g., the feedback

"you need to re-write this"). Furthermore, continued non-sense making will likely lead to aversive distal consequences (e.g., finding another line of work).

Consider another example of a husband who inadvertently offends his wife at dinner. The degree to which the husband provides a "sensible" and "reasonable" verbal explanation for his offensive behavior will likely directly influence the quality of interactions for the remainder of the evening. Over time, it is not difficult to see how repeated instances of "non-sensible" or "unreasonable" verbal explanations could negatively impact the overall stability and longevity of the relationship. These two examples highlight the adaptive role sense making plays in our social verbal community. The ability to engage in coherent verbal behavior appears to have direct implications for fulfilling humans' need to belong, both professionally and interpersonally (cf. Baumeister & Leary, 1995).

There is a wealth of evidence suggesting that sense making is adaptive on a psychological level. Mineka and Hendersen's (1985) review of predictability and control suggests that humans are motivated to make sense of their environment because doing so allows more effective action. In addition, sense making also appears to be central to the psychological construct of intelligence (IQ). Consider this contemporary and widely accepted definition of intelligence, "It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—'catching on,' 'making sense' of things, or 'figuring out' what to do" (Gottfredson, 1997, p. 13). Sense making is explicitly referenced in the definition of IQ and one of the most commonly used IQ tests, the Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV), directly assess sense-making behavior via a similarities subtest where individuals are asked to relate two words together (e.g., "how are a cup and a backpack alike?"; Wechsler, 2008). Given the well-established relationship between IQ

and successful life outcomes (see Kaufman & Lichtenberger, 2006 for a detailed review), it appears that sense making is a broadly adaptive behavior.

Sense making has also been shown to be adaptive following traumatic or psychologically difficult experiences. Expressive writing, in which individuals write detailed narratives of past negative events, has been shown to produce beneficial outcomes in individuals who have experienced traumatic events (Koopman et al., 2005; Smyth, Hockemeyer, & Tullock, 2008), a recent job loss (Spera, Buhrfeind, & Pennebaker, 1994), and depressive symptoms (Gortner, Rude, & Pennebaker, 2006; see Pennebaker & Chung, 2011 for a detailed review of the expressive writing paradigm). In reviewing mechanisms of change within the expressive writing paradigm, Pennebaker and Segal (1999) concluded that the formation of a coherent narrative is a critical component in generating positive psychological outcomes. Such findings are not unique to the expressive writing paradigm. For example, Mendola and colleagues (1990) found that forming positive causal attributions following impaired fertility was associated with improved psychological functioning. In addition, continued interest in narrative therapy techniques lends support to the positive effects of sense making in treating some psychological difficulties (Roberts, 2000). Taken as a whole, these findings suggest that sense making behavior leads to positive psychological outcomes in a variety of contexts.

Finally, studies investigating the aversive nature of ambiguity lend support to the adaptive nature of coherence. One interpretation of Wray's (2011) finding that participants overwhelmingly preferred a neutral or solvable context to an unsolvable one is that ambiguous situations are aversive. This interpretation is supported by other areas of psychological literature. For example, lack of actual or perceived control over the environment has been well established as a predictor of distress in both human and animal models (Maier & Seligman, 1976; see also

Burger & Arkin, 1980; Sanford, Yang, Wellman, Liu, & Tang, 2010).Ambiguity has also been directly shown to mediate physiological arousal of aversive events with increased skin-conductance and heart rate associated with ambiguous versus known aversive conditions (Sosnowski, 1983; 1988). Thus, it appears that coherence also serves a negative reinforcing function as making sense of the environment allows escape from the aversive condition of ambiguity.

  Sense making gone awry. Given the broad range of contexts where coherence appears to be adaptive it is tempting to draw the conclusion that making sense is ubiquitously adaptive. However, a growing body of evidence is emerging that suggests that sense making can be psychologically damaging in particular contexts. What follows, is a discussion of occasions in which making sense has been implicated in psychological distress.

  Early pioneers of cognitive therapy were among the first to explicate the potential dangers of sense making in a framework amenable to scientific testing. Beck (1976) developed a theoretical framework that posited that over-generalized negative thoughts about the self, world, and future were the primary maintaining variable of depression and other emotional disorders. Cognitive therapy's attention to core beliefs, defined as thoughts about the self that are global, rigid, and overgeneralized (Beck, 2011), drew scientific attention to the ways in which sense making can create and exasperate psychological difficulties. In particular, rumination, the behavior of trying to figure out either the reasons why something bad happened or the reasons for a negative emotional state, appears to be a prime example of the iatrogenic effects of sense making (Nolen-Hoeskema & Morrow, 1991; Smith & Alloy, 2009). Frequency and intensity of rumination has been repeatedly linked to increased depressed symptoms such as negative affect

and impaired cognitive functioning (Larsen, & Cowan, 1988; Lyubomirsky & Nolen-Hoeskema, 1995; Papageorgiou & Wells, 2003; Watkins & Moulds, 2005).

Rumination is of particular interest to the current investigation as findings have suggested that a large number of causal words are found in rumination writings (Smith & Alloy, 2009; Watkins, 2004). Additional research suggests that depressed individuals display poor autobiographical memories for specific events and instead rely on overgeneralizations and categorical thinking when discussing their past (Williams et al., 2007; Williams, Tesdale, Segal, & Soulsby, 2000). Making generalizations across situations and engaging in categorization are both examples of sense making behavior. Intervention research suggests that improving autobiographical memory among depressed individuals can lead to improvements in depressive symptoms (Neshat-Doost et al., 2012; Williams et al., 2000). These findings are suggestive of the iatrogenic role sense making behavior can play in depression.

The potential negative consequences of sense making are not limited to depression. For example, apophenia, the tendency to see patterns and meaning in randomness has been associated with increased levels of delusional thinking (Fyfe, Williams, Mason, & Pickup, 2008). This finding suggests that over-active sense making may play a role in psychotic disorders such as schizophrenia. The potential negative effects of sense making have also been considered by proponents of Acceptance and Commitment therapy (ACT; Hayes, Strosahl, & Wilson, 2011).

ACT is based on a relational frame theory (RFT) account of language and cognition and posits that much of psychological suffering can be attributed to verbal processes going awry (Wilson, Hayes, Gregg, & Zettle, 2001). Cognitive fusion, one of the primary sources purported to create and maintain psychopathology within the ACT model, speaks directly to the problematic nature of sense making. It is defined in ACT as the domination of verbal events (i.e.,

derived relational responding and sense making) in controlling responding to the exclusion of other contextual variables (Hayes et al. 2011, p. 69). Sense making behaviors such as engaging in excessive casual attributions, excessive attempts to figure out how to change, and excessive social comparisons are all examples of cognitive fusion. As ACT is a transdiagnostic model, cognitive fusion has been implicated as playing a role in a broad range of psychological disorders and problems in living (Blackledge, 2007). Specific conceptual attention to cognitive fusion has been given across a wide variety of domains such as depression (Kanter, Busch, Weeks, & Landes, 2008), anxiety (Friman, Hayes, & Wilson, 1998), parenting (Coyne & Wilson, 2004), and organizational behavior (Bond, Hayes, & Barnes-Holmes, 2006).

Little empirical consideration has been given to cognitive fusion itself as a pathological process. However, there is significant evidence that high levels of psychological inflexibility, a construct that includes both cognitive fusion and experiential avoidance (the degree to which an individual attempts to escape or avoid thoughts and feelings), are robustly associated with greater levels of psychological distress (see Ruiz, 2010 for a review; see also Bond et al., 2011; Hayes, Luoma, Bond, Masuda, & Lillis, 2006). In addition, the ACT intervention of defusion (i.e., therapeutic techniques aimed at reducing cognitive fusion) has routinely been shown to reduce the believability and discomfort caused by negative thoughts (Healy et al., 2010; Masuda, Hayes, Sackett, & Twohig, 2004; Masuda et al., 2009; Masuda et al., 2010). A recent meta-review of ACT component interventions found a significant medium effect size for defusion interventions across seven separate studies (Levin, Hildebrandt, Lillis, & Hayes, 2012). As a whole, the correlation evidence for psychological inflexibility and the component evidence for defusion suggest that excessive sense making may be a factor in a variety of psychological difficulties.

In addition, evidence for ACT as a treatment package also lends support for the role that verbal regulation and sense making play in psychological difficulties. ACT enjoys substantial evidence of efficacy across a variety of disorders and populations suggesting that defusion, one of the major treatment components, may be an active treatment component (see recent meta-reviews such as Hayes et al., 2006; Ost, 2008; Powers et al., 2009; Ruiz, 2012). More evidence of the non-adaptive consequences of sense making is found in mediational analyses of ACT interventions. For example, the display of verbal statements consistent with defusion during ACT therapy sessions has been shown to mediate clinically significant outcomes in a trial for tinnitus (Hesser, Westin, Hayes, & Andersson, 2009). In addition, a recent reanalysis of an ACT depression trial revealed that post treatment levels of cognitive defusion mediated follow up levels of depression (Zettle, Rains, & Hayes, 2011). These studies provide evidence that ACT works, at least in part, through the process of defusion providing further support for the casual role that excessive verbal regulation and sense making play in maintaining psychological difficulties. The ACT evidence taken as whole, and evidence for the process of defusion in particular, suggests that excessive verbal regulation and engagement in sense making are associated with a wide range of psychological difficulties.

Flexible sense making. A review of the evidence of positive and negative consequences of sense making yields ample examples of both. However, there does not appear to be clear agreement in delineating between contexts where sense making helps or hinders. That is, there appear to be contexts where existing evidence suggests that sense making can result in both adaptive and psychologically harmful consequences. For example, some studies have found sense making to be adaptive in addressing depressive symptoms while others have found the opposite to be true (Gortner, Rude, & Pennebaker, 2006; Papageorgiou & Wells, 2003). One

possible way in which to make sense of this discrepant finding is via the psychological flexibility model (Hayes, Strosahl, & Wilson, 2011; see also, Hayes, Levin, Plumb-Vliardaga, Vilatte, & Pistorello, 2012).

The psychological flexibility model is the overarching model that encompasses both acceptance and commitment therapy (ACT) and relational frame theory (RFT). The model consists of six core components that comprise psychological flexibility: cognitive defusion, acceptance, present moment focus, self as context, values, and committed action (Hayes et al., 2011). Within the model, psychological flexibility is defined as, "the ability to fully contact the present moment and the thoughts and feelings without needless defense, and, depending upon what the situation affords, persisting or changing in behavior in the pursuit of goals and values" (Bond et al., 2011, p. 8). Using this theoretical perspective, it is possible to view sense making behavior as being functional or dysfunctional depending on the contextual feature of the situations. That is, an individual high in psychological flexibility would be expected to successfully discriminate contexts where sense making behavior is useful and likewise let go of sense making behavior in contexts where it inhibits desired outcomes. In contrast, psychologically inflexible individuals would be expected to persist in sense making behavior even in contexts where such behavior is unlikely to help them achieve their desired goal. Thus, the psychological flexibility model provides a framework to predict how individuals will behave in tasks where sense making behavior becomes increasingly less adaptive.

Current Study

This study will explore both basic properties and applied implications of coherence. With regard to basic properties, this investigation seeks to extend the literature by testing the basic assumption that coherence is both a well-established repertoire and a potential reinforcer in

verbally competent humans. In particular, this study seeks to determine whether or not participants will demonstrate spontaneous coherent responding in the absence of direct contingencies programmed by the experimenter. To achieve this aim, participants will be randomized to receive different learning histories with arbitrary experimental stimuli. It is hypothesized that slight differences in learning history will result in different patterns of responding on an ambiguous task. It is also hypothesized that participants will report self-generated verbal rules that are largely consistent with their response patterns when asked to do so. That is, when asked to verbally make sense they will report using strategies that are consistent with their previous responding. Furthermore, it is predicted that after reporting self-generated verbal rules participants will behave more consistently with their own rules when asked to complete an ambiguous task for a second time.

In addition to assessing for the spontaneous display of coherent responding, this study seeks to replicate and extend the work of Wray (2011) by evaluating the extent to which verbally competent humans show a preference towards coherent responding. It is hypothesized that when given a choice between contexts where coherent responding is possible or impossible, participants will display a preference towards the coherent context. It is also anticipated that participants will, to varying degrees, persist in preference towards coherent contexts even when an aversive consequence is in place. The extent to which participants persist in preference towards coherence in the face of an aversive is expected to provide a measure of the relative strength of each participant's coherent response repertoire and the degree to which coherence may function as a reinforcer.

This study also seeks to explore the applied implications of coherence by examining the relationships between responding on self-report measures of psychological flexibility, cognitive

flexibility, and psychological distress, and the degree of persistence in preference towards coherent contexts. In particular, it is hypothesized that participants with high levels of psychological inflexibility and cognitive fusion will show greater persistence in preference towards coherent contexts than participants with low levels of these constructs. Furthermore, it is hypothesized that participants who self-report higher levels of psychological distress will show greater persistence in preference toward coherent contexts. Finally, it is hypothesized that there will be orderly differences in participants' self-reported level of frustration throughout the experimental task. In particular, it is anticipated that psychological inflexibility and cognitive fusion will moderate self-reported levels of frustration with participants who report high psychological inflexibility and high cognitive fusion displaying higher levels of frustration throughout the experiment.

METHODS

Participants and Setting

Undergraduate students who were enrolled in a psychology department undergraduate subject pool at a large southern university served as participants in this experiment. Subjects received one hour of experimental credit in exchange for their participation.

The experiment was presented on a Dell desktop computer running Windows XP Service Pack 2. Participants were seated at a desk in front of a 17-inch color computer monitor in one of two 8' by 10' rooms. Only one participant was run in each experimental room at a time and participants were left alone to complete the experiment. Participants completed initial self-report measures using the browser based Qualtrics survey system. Responses on the initial self-report measures were stored on a secured university owned server. The remainder of the experimental paradigm was programmed in Visual Basic 2008 Professional Edition and was compiled into a stand-alone executable program to ensure consistency across participants. Participant responses were written directly by the program to a Microsoft Access database to ensure accurate recording of all study variables.

Measures

General Health Questionnaire (GHQ-12). The GHQ-12 is a 12-item four-response option scale designed to measure psychiatric morbidity in general practice settings (Goldberg, 1978). While originally developed in the 1960's as a 60-item scale (Goldberg & Blackwell, 1970), the 12 item short form of the instrument has been shown to have excellent psychometric

properties equivalent to or better than the properties of the original 60-item version (Goldberg et al., 1997). For example, a multi-site evaluation of the GHQ-12 revealed an average area under the receiver operating characteristic (ROC) curve of .88 (sensitivity 83.4%, specificity 76.3%) when using the instrument as a screener to predict the presence of a ICD-10 or DSM-IV mental disorder (Goldberg et al., 1997). The GHQ-12 also enjoys solid reliability with scale alpha's ranging from .78 to .95 across evaluation studies (Jackson, 2007). While several different subscales and factor structures for the GHQ-12 have been proposed (Kalliath, O'Driscoll, & Brough, 2004), little utility has been found in using sub-scales for diagnostic prediction (Cleary, Goldberg, Kessler, & Nycz, 1982). The instrument is scored using a binary 0-0-1-1 scoring method such that the two most symptomatic responses are scored one and the two least symptomatic responses are scored zero for each item (Goldberg, 1978). A cut-off score of 2/3 has been identified as the threshold that yields the greatest specificity and sensitivity as a case predictor (Jackson, 2007; Makowska, Merecz, Moscicka, & Kolasa, 2002). In the current study sample, the internal consistency of the GHQ-12 was $\alpha = .86$, indicating good internal consistency.

      Acceptance and Action Questionnaire-II (AAQ-II). The AAQ-II is a seven-item seven point likert scale that measures psychological inflexibility (Bond et al., 2011). Scores on the unidimensional measure range from 7 to 49 with higher scores indicating greater levels of psychological inflexibility. While a relatively new psychometric instrument, the AAQ-II has good internal consistency (mean $\alpha = .84$, range of $\alpha = .78$ to $\alpha = .88$ across six validation samples with a total N of 2,816) and good temporal stability (one year test retest $r = .79$ in a sample of 583 British financial workers; Bond et al., 2011). The AAQ-II shows solid convergent validity with measures of psychological wellbeing (e.g., depression and anxiety) with greater levels of psychological flexibility associated with greater psychological wellbeing (Ruiz, 2010). In

addition, the measure has solid predictive validity of behavioral health outcomes with greater

psychological flexibility associated with fewer work absences and greater work productivity

(Bond et al., 2011). A recent item response theory based evaluation of the AAQ-II has

confirmed the unidimensional factor structure of the AAQ-II (Fledderus, Voshaar, Ten Klooster,

& Bohlmeijer, 2012).   In addition, the study also demonstrated that the measure possesses

incremental validity beyond established measures of mindfulness in predicting indicators of

positive mental health. In the current study sample, the internal consistency of the AAQ-II was α

= .88, indicating good internal consistency.

Cognitive Fusion Questionnaire (CFQ).   The CFQ is a 13-item seven point likert scale

that measures the construct of cognitive fusion. Scores range from 13 to 91 with higher scores

indicating greater levels of cognitive fusion. Cognitive fusion is a central target of intervention

within the psychological flexibility model and is defined as, "the extent to which we are

psychologically entangled with and dominated by the form or content of our thoughts (Gillanders

et al., under review, p. 3). The scale was developed and refined as part of two doctoral

dissertation projects (Dempster, 2009; Campbell, 2010). In its current form, the CFQ enjoys

solid evidence of internal consistency with a total scale alpha = .84 and a unifactorial structure as

confirmed by several independent confirmatory factor analyses (Gillanders et al., under review).

The CFQ also enjoys good temporal stability as evidenced by a four-week test-retest r = .82 in a

sample of 88 British prison service officers (Gillanders et al., under review). The CFQ enjoys

solid evidence of discriminant validity in differentiating between adults with and without

psychological disorders. In addition, the CFQ has solid initial evidence of convergent validity

with higher scores on the measure associated with greater psychological distress and

psychological inflexibility (Gillanders et al., under review). In the current study sample, the internal consistency of the CFQ was α = .64, indicating questionable internal consistency.

Marlowe-Crowne Social Desirability Scale- Short Form C (MCSDS-SF). The MCSDS-SF is a 13-item true/false response scale that measures social desirability. In its original form, the MCSDS contained 33 true/false response items that assess the tendency of an individual to present a superlative picture of themselves by endorsing uncommonly possessed positive traits (e.g., "I'm always willing to admit it when I make a mistake") and failing to endorse commonly possessed negative traits (e.g., "I sometimes feel resentful when I don't get my way;" Crowne & Marlowe, 1960). The Reynolds (1982) 13-item short form (MCSDS-SF) displays adequate internal consistency (α = .76) when compared to the full 33 item scale (α = .82) and good convergent validity with the full 33 item scale (r=.93). Thus, the MCSDS-SF appears to maintain the solid psychometric properties of the 33 item original scale while gaining the advantage of brevity (Reynolds, 1983). Scores on the MCSDS-SF range from 0-13 with greater scores indicating greater presence of a socially desirable response tendency. The MCSDS-SF is included in the current study following Maher's (1978) recommendation that the response tendency of social desirability be assessed and controlled for in psychological research. In the current study sample, the internal consistency of the MCSDS-SF was α = .70, indicating acceptable internal consistency.

Demographic Survey. Participants were asked to report their age, gender, racial and ethnic identity, and current year in school.

Frustration Visual Analogue Scale (F-VAS). A computer based visual analogue scale was used to assess participants' subjective rating of frustration. The written prompt for the F-VAS consisted of the question "How frustrated do you feel right now?" with the lower bound

anchor consisting of the words "Not at all" and the upper bound anchor consisting of the word "Extremely." Participants used a computer mouse to drag a slider along the 160mm wide scale to indicate their current level of frustration. The computer program converted the position of the slider into a value between 0 and 100 with 0 indicating no frustration and 100 indicating extreme frustration. The slider was placed at the halfway point (i.e., 50) during each F-VAS presentation and participants were required to click and move the slider prior to registering their response. The use of pen and paper based visual analogue scales have been well established as measures of subjective psychological constructs such as mood (Cella & Perry, 1986) and pain (Price, McGarth, Rafili, Buckingham, 1983). In addition, a recent empirical investigation has demonstrated that a computer based visual analogue scale perform equally well in terms of reliability and validity as its pen and paper counterpart (Kreindler, Levitt, Woolridge, & Lumsden, 2003).

Design

This study utilized both within-subject and between-subject design elements to assess study hypotheses. Given the complexity of the design, the study was conducted in six phases with all participants flowing through each phase in sequence. A brief description and rationale of each phase is provided below.

Initial self-report. Participants completed a series of psychometrically validated self-report measures. In addition, they rated their current level of frustration on a visual analogue scale. The self-report measures were used to explore the ways in which the constructs they measure relate to performance during other study phases.

Practice. Participants were given brief written instructions and then exposed to a practice matching to sample preparation where they were asked to match names of colors (i.e.,

the word "red") to pictures of colors (i.e., a red square). This phase was designed to introduce

participants to the matching to sample task and allow them practice with the interface prior to the

experimental manipulation.

Coherence testing. Participants were randomly assigned to receive or not to receive a

task where they relate pictures of foods to concepts (i.e., healthy, unhealthy, and disgusting). All

participants were then asked to relate pictures of foods to arbitrary shapes. Performance on the

shapes task was analyzed to determine whether or not the different learning histories produce

different patterns of responding. During this task, participants were also asked to self-report the

meanings of the arbitrary shapes and the rule(s) they used to guide their responding. These self-

reported rules were used to determine whether or not participants followed the rules they stated

and whether or not the self-reporting of rules lead to more rule-consistent behavior.

Class acquisition.   Participants were exposed to a matching to sample training paradigm

where they were given corrective feedback while relating nonsense syllables and shapes together.

After meeting a standardized performance benchmark, participants were tested for the display of

combinatorial entailment (i.e., equivalence relations) between experimental stimuli. This phase

provided the learning history necessary for participants to interact meaningfully with the stimuli

used during the final two phases of the study.

Coherence preference assessment.   Participants were introduced to a computerized

concurrent chain schedule procedure where they were asked to select between matching to

sample trials that were consistent with their learning history in the class acquisition phase (i.e.,

coherent trials) or trials that were inconsistent with their learning history in the class acquisition

phase (i.e., incoherent trials). One color was associated with coherent trials and another color

was associated with incoherent trials during this phase. This phase allowed for the direct

assessment of participant preference towards coherent contexts. Participants were also asked to self-report the meaning of each color halfway through this phase to allow for the assessment of whether or not self-reporting of rules lead to more rule-consistent behavior.

Coherence preference assessment with response cost. Participants were exposed to an identical procedure as in the previous phase except that a response cost in the form of an increased delay between trials was added to the coherent response option. The inter-trial interval (ITI; i.e., delay between trials) for the coherent response option was systemically increased from two seconds to seven seconds across 90 trials while the ITI for the incoherent response option remained at one second throughout. This phase allowed for the assessment of the degree to which participants persist in preference towards coherent contexts when faced with an increasingly aversive consequence.

Stimuli

Stimuli used in practice phase. Two three-member stimulus classes (W and X) were used during the practice phase of the study. These practice stimuli were designed to allow participants to gain exposure to the matching to sample procedure prior to use of the experimental stimuli. The practice stimuli are presented in Figure 2. The W class consisted of 4" x 4" images of colors (red, green, and black) and the X class consisted of 4" by 4" images of the words "Red", "Green", and "Black" in 48 point Arial Black font on a white background.
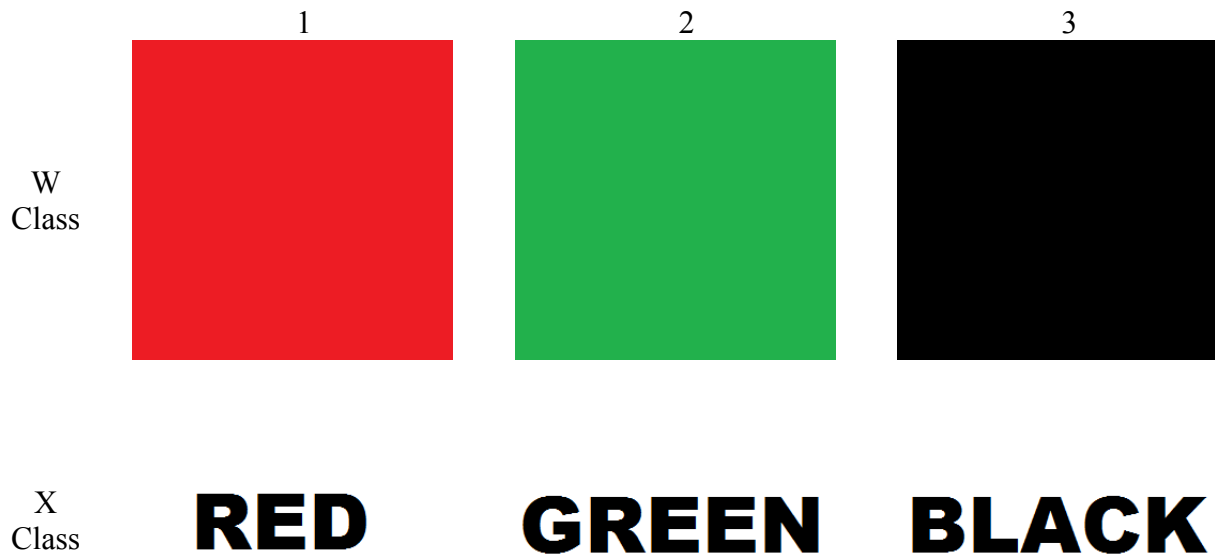
Figure 2. Stimuli used during the practice phase of the study.

Stimuli used in coherence testing phase. Two three-member stimulus classes (Meaning and Shapes) and one 27-member stimulus class (Foods) were used for the coherence-testing phase of the study. The Meaning and Shapes class members are presented in Figure 3. The Meaning class consisted of 4" by 4" images of the words "Healthy", "Unhealthy," and "Disgusting" in 36 point Arial font on a white background. The Shapes class consisted of 4" by 4" images of shapes (lines, circles, and open triangles) drawn by the experimenter in Microsoft Paint.

	1	2	3

Meaning
Class   Healthy   Unhealthy   Disgusting
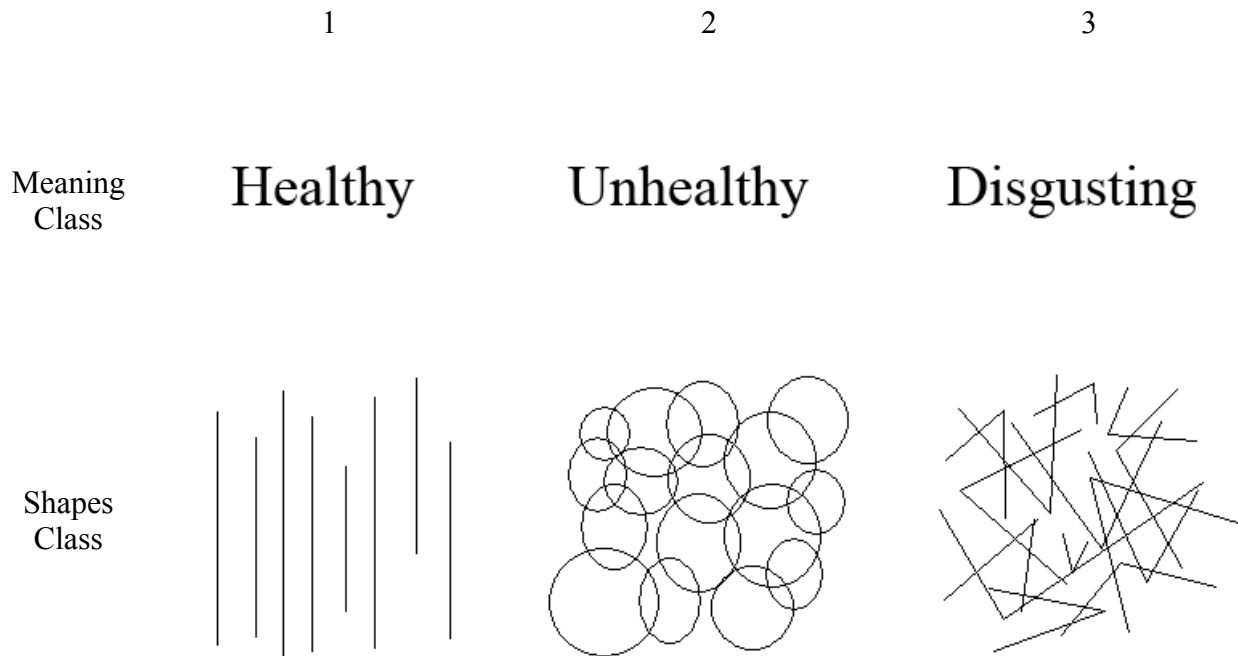
Shapes
Class

Figure 3. Meaning and Shapes stimuli used during the coherence testing phase of the study.

The 27 members of the Foods class consisted of 4" by 4" images of various foods and are presented in Figure 4. The images were acquired from the Internet with Google image search used to locate appropriately sized images. The Foods class was designed to allow for the images to be sorted into two distinct subclasses defined by either topographical appearance (i.e. shape) or the functional property of healthiness (i.e., meaning). Each Foods class stimulus was designed to belong both to a specific shape (i.e., line, circle, or triangle) and meaning (i.e., healthy, unhealthy, or disgusting) subclass. The subclasses were balanced such that the 27-member class could be sorted completely by shape (nine class members each belonging to line, circle, and triangle, respectively) or meaning (nine class members each belonging to healthy, unhealthy, and disgusting, respectively).

	Healthy	Unhealthy	Disgusting
Lines			
Circles			
Triangles			

Figure 4. Food stimuli used during the coherence testing phase of the study. The 27 members of the Foods class were designed to be sorted into two distinct subclasses. The meaning subclass was based on the functional property of healthiness and is presented in the figure vertically with nine Foods class members each belonging to the subclass categories of healthy, unhealthy, and

disgusting, respectively. The shapes subclass was based on topographical appearance and is presented in the figure horizontally with nine Foods class members each belonging to the subclass categories of lines, circles, and triangles.

Stimuli used in class acquisition phase. Three 3-member stimulus classes (A, B, and C) were used during the class acquisition phase of the study. All nine stimuli used during the class acquisition phase are presented in Figure 5. Both the A and B classes were comprised of 4" by 4" images of unique, randomly generated nonsense syllables (e.g., LOD) in 72 point Arial Black font on a white background. The C class was comprised of 4" by 4" images of novel drawings made by the experimenter in Microsoft Paint.
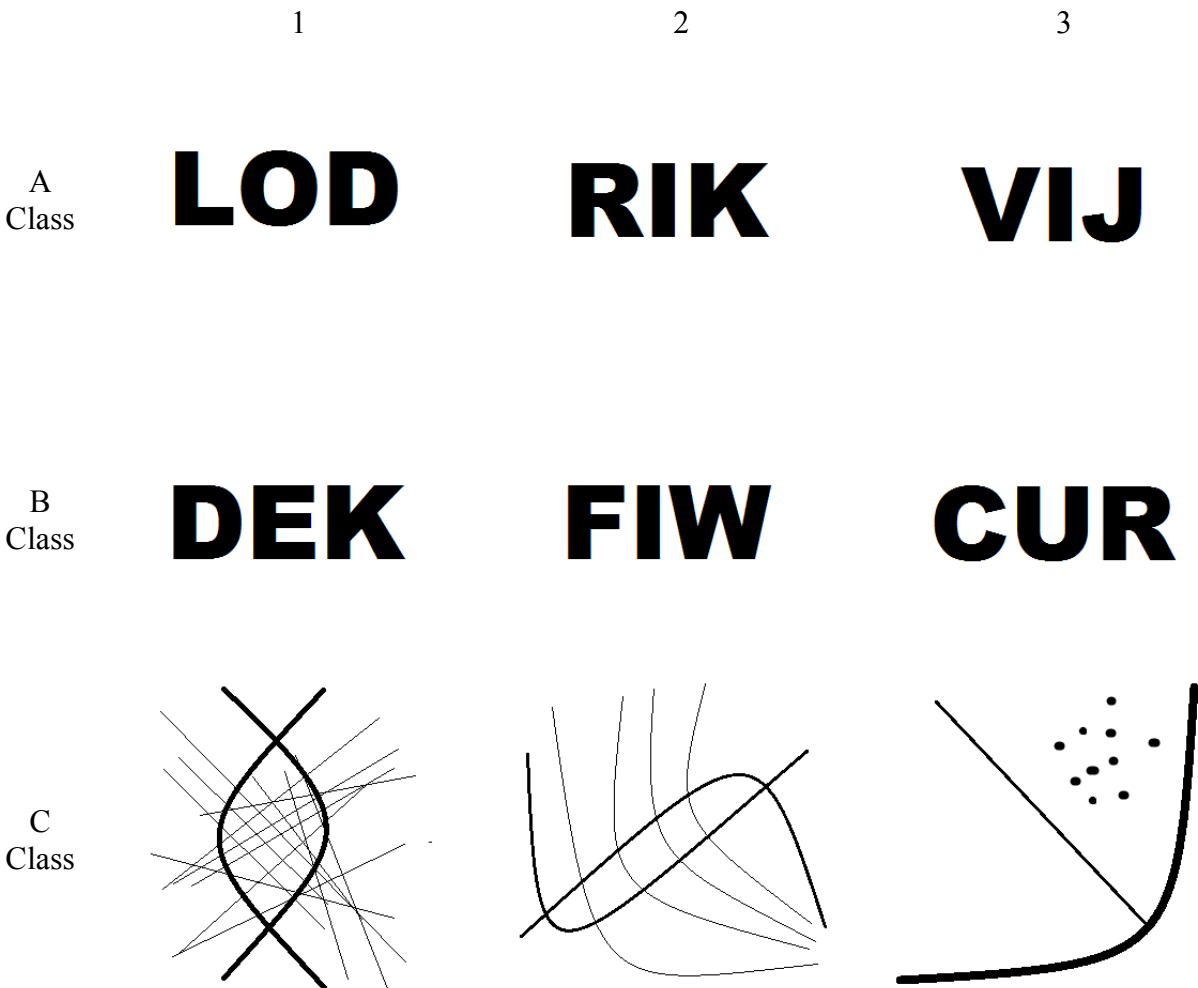
Figure 5. Stimuli used during the class acquisition phase of the study.

Stimuli used in coherence preference assessment & coherence preference assessment with response cost phases. Two 4" by 4" images of colors (blue and yellow) were used as contextual cues for the initial link of the concurrent chain schedules in both the coherence preference assessment and coherence preference assessment with response cost phase of the study. The images used for the contextual cues are presented in Figure 6.
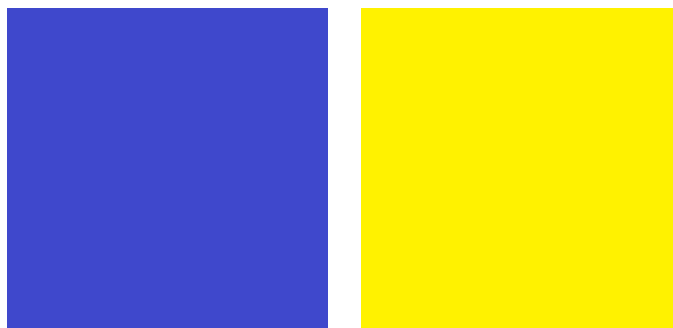
Figure 6. Stimuli used for contextual cues for the initial link of the concurrent chain schedules.

Procedure

Upon presenting to the experimental setting participants were asked for informed consent and given the opportunity to ask questions about the study. Participants were provided with the following description of the study during informed consent:

> We are interested in investigating the ways in which people make sense of ambiguous tasks. In order to reach this aim, we are asking you to fill out a series of questionnaires and complete some computer tasks that require you to relate items together in different ways. The experiment will take approximately one hour to complete in our laboratory.

In addition, to comply with Institutional Review Board (IRB) guidelines regarding risk disclosure, participants were informed that, "you may feel frustrated during this study as you try to make sense of ambiguous tasks." After informed consent was obtained, participants began the experimental sequence with the initial self-report phase.

Initial self-report. Upon being introduced to the computerized experimental interface, participants were asked to complete the following self-report measures: GHQ-12, AAQ-II, CFQ, MCSDS-SF, and the demographic survey. Participants were then asked to report their current level of frustration using the F-VAS. This was the first of 17 occasions during the course of the study where participants were asked to rate their frustration level using the F-VAS. The locations of all F-VAS administrations within the experimental sequence are presented in Table 1. For the sake of clarity and brevity, future occurrences of F-VAS administrations will be omitted from this manuscript for the remainder of the procedure section.

Table 1.

Location of all F-VAS Administrations Within the Experimental Sequence

Study Phase	Location (immediately following)
Initial Self-Report	
F-VAS 1	Demographic survey
Practice	
F-VAS 2	Last trial of W-X testing block
Coherence Testing	
F-VAS 3	Last trial of first testing block
F-VAS 4	Last trial of second testing block
Class Acquisition	
F-VAS 5	Last trial of A-B training block[a]
F-VAS 6	Last trial of A-C training block[a]
F-VAS 7	Last trial of mixed A-B/A-C training block[a]
F-VAS 8	Last trial of B-A/C-A testing block
F-VAS 9	Last trial of B-C/C-B testing block
Coherence Preference Assessment	
F-VAS 10	Last trial of pre-rule equal block
F-VAS 11	Last trial of post-rule equal
Coherence Preference Assessment with Response Cost	
F-VAS 12	Last trial of +1 second response cost block
F-VAS 13	Last trial of +2 second response cost block
F-VAS 14	Last trial of +3 second response cost block
F-VAS 15	Last trial of +4 second response cost block
F-VAS 16	Last trial of +5 second response cost block
F-VAS 17	Last trial of +6 second response cost block

[a]When 89% correct response criterion is met

Practice. At the beginning of this phase participants were exposed to a screen with the following written instructions, "When the experiment begins, images will appear on the computer screen. Your task is to choose one image from the options on the lower portion of the screen. Click continue when you are ready." After reading the instructions and asking any questions, participants clicked on a continue button to begin the first phase of the experiment.

An arbitrary, simultaneous matching to sample procedure (Green and Saunders, 1998) was used to test for relational discriminations among the two practice stimuli classes (W and X). During each trial, a sample stimulus was presented in the upper middle portion of the screen. In addition to the sample stimuli, three comparison stimuli were presented in the lower left, lower middle, and lower right portions of the screen. A screen shot of the matching to sample interface is provided in Figure 7. The three comparison stimuli were randomly positioned throughout the study to control for a position based response bias. In addition, the order of the trials within each testing and training block were randomized for each participant to control for possible intra-block order and sequence effects. Participants selected among the comparison stimuli by clicking on one of the three comparison stimulus with the computer mouse. Following each trial there was a one second inter-trial-interval (ITI) in which no programmed reinforcement was provided during this phase.
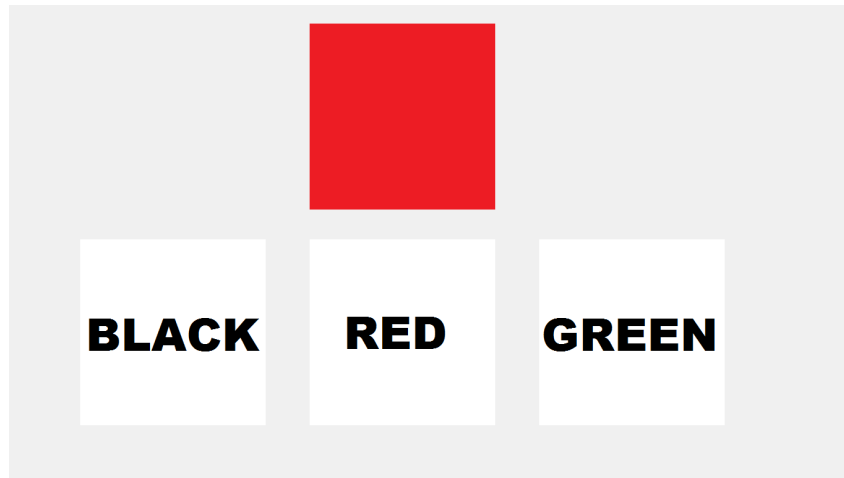
Figure 7. Screen shot of the matching to sample interface.

During the practice phase participants were exposed to one testing block which assessed the following stimulus relations; W1-X1, W2-X2, W3-X3, X1-W1, X2-W2, and X3-W3. Each relation was tested three times for a total of 18 trials. As this phase was primarily designed to introduce participants to the matching to sample interface, there was no performance requirement for participants to advance to the next phase of the study.

Coherence testing. An identical matching to sample procedure as used in the practice phase of the study was used to test for relational discriminations among coherence testing phase stimuli. Participants were not provided with any written instructions at the beginning of this phase. Prior to beginning this phase each participant was randomized to either the Meaning or Shapes condition via the use of a computerized random number generator.

Participants assigned to the Meaning condition were initially exposed to a 27 trial block that required them to sort the 27 Foods class stimuli using the three members of the Meaning class (i.e., Healthy, Unhealthy, and Disgusting) as comparison stimuli. No programmed reinforcement was provided during this block and following completion of the block Meaning condition participants proceeded to the first testing block. Participants assigned to the Shapes

condition were not exposed to the meaning testing block and instead proceed directly to the first testing block. Thus, the only difference between conditions is that participants assigned to the Meaning condition had the additional learning history of using the Meaning class stimuli to sort the Foods class stimuli without any programmed reinforcement.

During the first testing block all participants were asked to sort the 27 Foods class stimuli using the three members of the Shapes class (i.e., lines, circles, and triangles) as comparison stimuli without any programmed reinforcement. Immediately following the first testing block participants were given the following instructions, "In the space below, please write down the rule(s) you used to match the symbols to the pictures." Participants typed their responses using the computer keyboard and clicked a button to continue when they were ready. Following the open ended self-report task, participants were exposed individually to each of the three members of the Shapes class along with the written instructions, "What does the symbol above mean?" Participants typed their response for each Shapes class member and then click a button to continue. Immediately following the self-report tasks, participants were exposed to a second testing block that was identical to the first testing block. That is, all participants once again were asked to sort the 27 Foods class stimuli using the three members of the Shapes class without any programmed reinforcement. Following the completion of the second testing block participants advanced to the next phase of the study.

Class acquisition. During this phase of the study, participants were exposed to an established sequence of matching to sample procedures to train and subsequently test for the derivation of stimulus relations among three 3-member classes (cf. Steel & Hayes, 1991, and Green & Saunders, 1998).   In particular, participants were sequenced through three training blocks (A-B, A-C, and mixed A-B / A-C) and two testing blocks (B-A / C-A, and B-C / C-B).

41

An overview of the number of trials and response criterion for each block is provided in Table 2, and an explication of all the trained and tested relations by block is provided in Table 3.

Table 2.

Number of Trials and Response Criterion for Each Block During the Class Acquisition Phase

Block	Trials	Training Criterion	Testing Criterion
Train A-B	18	16/18 (89%)	
Train A-C	18	16/18 (89%)	
Mixed Train A-B and A-C	36	32/36 (89%)	
Test B-A and C-A	18		16/18 (89%)
Test B-C and C-B	18		16/18 (89%)

Table 3.

Trained and Tested Relationships by Block During the Class Acquisition Phase

Block	Trained Relationships	Tested Relationships
Train A-B	A1-B1 A2-B2 A3-B3	
Train A-C	A1-C1 A2-C2 A3-B3	
Mixed Train A-B and A-C	A1-B1 A1-C1 A2-B2 A2-C2 A3-B3 A3-C3	
Test (Mutual Entailment) B-A and C-A		B1-A1 C1-A1 B2-A2 C2-A2 B3-A3 C3-A3
Test (Combinatorial Entailment) B-C and C-B		B1-C1 C1-B1 B2-C2 C2-B2 B3-C3 C3-B3

Participants began this phase in the A-B training block where they were exposed to six training trials for each A-B relation for a total of 18 training trials. Matching to sample procedures identical to those in earlier study phases will be used with the exception that corrective feedback was issued during the one-second inter-trial-interval immediately following each trial. More specifically, the word "Correct" appeared following accurate conditional discriminations and the word "Incorrect" appeared following inaccurate conditional discriminations. Participants cycled through A-B training blocks until they reach the required criterion of 16 out of 18 correct responses.

The A-C training block consisted of identical procedures as the A-B block with the only difference being the relations trained. After meeting the response criterion of the A-C block, participants were exposed to a 36-trial mixed training block consisting of 18 A-B training trials and 18 A-C training trials. Once participants met the required 32 correct response criterion they were sequenced to the testing trial blocks.

Participants were first tested for the mutual entailment of stimulus relations in the B-A / C-A testing block. Identical procedures were used in this testing block as in training except that corrective feedback was no longer provided during inter-trial-intervals. That is, no programmed reinforcement was provided during testing trials. Participants were exposed to a total of 18 testing trials during this block and were sequenced to the next block regardless of the number of correct responses. However, participants were considered to have displayed mutual entailment only if they emit 16 or more correct responses during this testing phase. During the B-C /C-B testing block combinatorial entailment of stimulus relations was assessed using identical testing procedures as the previous testing block. Participants advanced to the next phase of the study regardless of performance in this phase; however, participants were considered to have displayed combinatorial entailment only if they emit 16 or more correct responses during this testing phase.

Coherence preference assessment. A modified concurrent chains schedule procedure was used during this phase to assess participant response preferences towards coherent and incoherent matching to sample trials. Figure 8 provides a graphical overview of the modified concurrent chain procedure. During the initial link of the chain, participants were exposed to two concurrently available stimuli consisting of a 4 " by 4" blue square and a 4" by 4" yellow square. The position of each stimulus was randomized during each presentation such that they either appeared in the middle left or middle right of the screen. This randomization was

designed to control for a position response bias and ensured that the participants responded to the relevant feature of the stimuli (i.e., color) instead of position.

Access to the mutually exclusive terminal links was available via a single response (i.e., mouse click) on one of the initial link stimuli. That is, both initial link schedules provided access on a fixed response one (FR1) schedule to their respective terminal link stimuli. Both terminal link stimuli consisted of a standard matching to sample trial presentation comprised of stimuli from the class acquisition phase of the study. The only difference between the terminal link stimuli was that one of the terminal links consisted of matching to sample trials that were consistent with the training provided during the class acquisition phase. The other terminal link consisted of matching to sample trials that were deliberately inconsistent with the participants' learning history during the class acquisition phase of the study. That is, one terminal link allowed access for coherent responding while the other link presented stimuli in such a way that coherent and internally consistent responding was impossible. The possible stimuli configurations used during coherent terminal links and incoherent terminal links are provided in Tables 4 and 5, respectively.
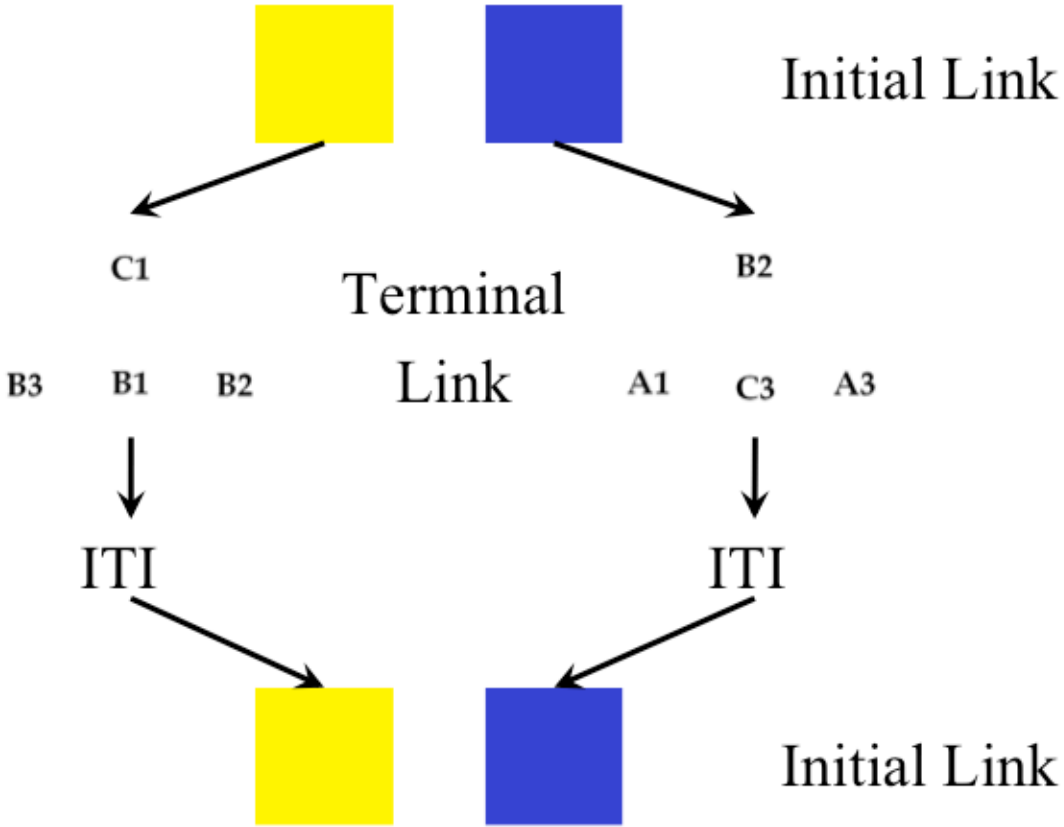
Figure 8. Graphical overview of the modified concurrent chain procedure. In this overview, the yellow initial link is connected to the coherent terminal link and the blue initial link is connected to the incoherent terminal link.

Table 4.

Stimuli Configurations Possible During Coherent Terminal Links

Sample	Comparison	Sample	Comparison
B1	A1, A2, A3	B1	C1, C2, C3
B2	A2, A1, A3	B2	C2, C1, C3
B3	A3, A1, A2	B3	C3, C1, C3
C1	A1, A2, A3	C1	B1, B2, B3
C2	A2, A1, A3	C2	B2, B1, B3
C3	A3, A1, A2	C3	B3, B1, B2

Note: During each coherent terminal link one of the 12 trial presentations shown above was randomly selected and displayed to the participant.

Table 5.

Stimuli Configurations Possible During Incoherent Terminal Links

Sample	Comparison	Sample	Comparison
B1	C2, A2, A3	B1	A2, C2, C3
B1	C3, A2, A3	B1	A3, C2, C3
B2	C1, A1, A3	B2	A1, C1, C3
B2	C3, A1, A3	B2	A3, C1, C3
B3	C1, A1, A2	B3	A1, C1, C2
B3	C2, A1, A2	B3	A2, C1, C2
C1	B2, A2, A3	C1	A2, B2, B3
C1	B3, A2, A3	C1	A3, B2, B3
C2	B1, A1, A3	C2	A1, B1, B3
C2	B3, A1, A3	C2	A3, B1, B3
C3	B1, A1, A2	C3	A1, B1, B2
C3	B2, A1, A2	C3	A2, B1, B2

Note: During each incoherent terminal link one of the 24 trial presentations shown above was randomly selected and displayed to the participant.

The association between initial and terminal link stimuli was randomized for each

participant. More specifically, for approximately half of the participants yellow linked to the

coherent terminal link and blue linked to the incoherent terminal link and for the other half of

participants the association was switched. No programmed reinforcement was provided

following the selection of a comparison stimuli on the terminal link. Responses on both terminal

links resulted in a one second inter trial interval (ITI) during which nothing was displayed on the

screen. At the end of the ITI participants were cycled back to the concurrently available initial

link stimuli.

Participants were provided with the following written instructions at the beginning of this

phase.

> In the next part of the study there is a blue and yellow option. Sometimes you will be
> forced to pick one and other times you will be able to choose for yourself. Please take a
> short break and when you are ready to continue click the button below.

Upon clicking continue, participants were exposed to ten forced choice trials whereby only one of the initial link stimuli appeared on the screen at a time. This forced choice procedure was designed to ensure that participants had adequate contact with both terminal link conditions. In addition, the forced choice trials allowed for participant responding to come under the stimulus control of the initial link stimuli such that the initial link associated with the coherent terminal link could become a contextual cue for coherence. Likewise, the initial link associated with the incoherent terminal link could become a contextual cue for incoherence.

Following the 10 forced choice trials, participants were exposed to 30 concurrent choice trials where they were asked to choose between the initial link stimuli at the beginning of each trial. At the end of the 30th concurrent choice trial participants were given the following instructions, "In the space below, please write down the rule(s) you used to pick a color." Participants typed their responses using the computer keyboard and clicked a button to continue when they are ready. Following the open ended self-report task, participants were exposed individually to each of two initial link stimuli along with the written instructions, "What does the color above mean?" Participants typed their response for each initial link stimuli and then clicked a button to continue. After the self-report task, participants were exposed to 10 additional forced choice trials followed by another block of 30 concurrent choice trials.

Coherence preference assessment with response cost. The same concurrent chains procedure as used in previous phase was used during this phase with the following exceptions. This phase of the study was designed to assess strength of preference towards the coherent terminal link by introducing a response cost in the form of a gradually increasing time delay for only the coherent terminal link. That is, the length of the coherent terminal link ITI gradually increase while the incoherent terminal link ITI was held constant at one second. It was expected

47

that longer ITIs would have an aversive function for participants who typically want to get through a study as quickly as possible. The number of total trials remaining was displayed to participants during each initial link in an effort to increase the likelihood that increased ITI duration would have an aversive function. In addition, a countdown timer that shows the number of seconds remaining in the ITI was displayed to participants during each ITI throughout this phase. This timer was added to increase the salience of the differences in ITIs between coherent and incoherent terminal links. As with the previous phase, no programmed reinforcement was provided for responses during the terminal links.

Participants were exposed to six concurrent chain blocks each consisting of four initial forced choice trials followed by 15 concurrent choice trials. The ITI of the coherent terminal link was parametrically increased by one second across the six blocks. The coherent terminal link began with one second of additional delay relative to the incoherent terminal link culminated with six seconds of additional delay relative to the incoherent terminal link during the sixth block. The ITI of the incoherent terminal link remained constant at one second across all six blocks. An overview of the ITIs across the six blocks is presented in Table 6. Following the completion of the last block, participants were given a debriefing form, thanked for their time, and dismissed from the study.

Table 6.

Inter Trial Intervals (ITIs) for Each Terminal Link Across the Six Blocks of the Coherence Preference Assessment with Response Cost Phase.

Block	Coherent ITI (seconds)	Incoherent ITI (seconds)	Relative Difference (seconds)
+1 second response cost	2	1	+1 coherent
+2 second response cost	3	1	+2 coherent
+3 second response cost	4	1	+3 coherent
+4 second response cost	5	1	+4 coherent
+5 second response cost	6	1	+5 coherent
+6 second response cost	7	1	+6 coherent

RESULTS

Analytic Strategy and Data Screening

A combination of within subject and between subject analytic strategies were used to assess study hypotheses. When possible, techniques from both analytic traditions were utilized to provide a more comprehensive accounting of participant responding. Given the complexity of this analytic strategy, particulars of each analysis will be provided immediately before their respective findings.

Prior to the evaluation of study hypotheses, the dataset was screened for accuracy of values, missing data, and the fit between the obtained data and the assumptions of parametric statistical analysis (i.e., normality, linearity, homoscedasticity, and non-multicollinearity/ singularity). Data from eight participants (8.99% of the original sample of 89) were excluded due to missing data caused by a programming error (n = 7) or experimenter error (n = 1). Four of these excluded participants were assigned to the Meaning condition, three to the Shapes condition, and one was exposed to both conditions due to experimenter error. Data from the remaining 81 participants were valid, in range, and complete with no missing values.

A univariate and multivariate outlier analysis was then conducted to identify normatively extreme values within the dataset. Three cases (3.70%) were identified as containing univariate outlying values ($z \geq 3.29$, $p < .001$, two-tailed test) on study variables. One case had an age (54 years old) that was substantially greater than the sample distribution, $z = 8.80$. The case was

49

retained, as the study hypotheses were not expected to be moderated by age of adult subjects. One case had a practice test score (66%) that was substantially lower than the sample distribution, z = -7.3. A response pattern analysis of the participant's performance on the color matching practice test provided no conclusive evidence that the participant had difficulty making discriminations between the colors (i.e., color blindness). Instead, it suggested that the participant simply made several random errors during the task and the case was retained. The final outlying case had a frustration rating during the practice task (F-VAS 2 = 100) that was substantially greater than the sample distribution, z = 3.53. The case was retained, as none of the participant's 16 other F-VAS values were identified as extreme outliers. An analysis of multivariate outliers found no cases with a Mahalanobis distance exceeding the critical value of 67.985 (α = .001) or a leverage value exceeding the critical value of .862 (α = .001). Thus, all 81 cases were retained for analysis.

Several study variables displayed a departure from normality. Namely, some time points of the frustration visual analogue scale (F-VAS) approximated a bimodal distribution. However, parametric tests, such as the ANOVA model planned to analyze F-VAS scores, are robust to violations of normality with sufficient sample size and degrees of freedom (Tabachnick & Fidel, 2007). This dataset was of sufficient size to permit parametric analysis of variables even in cases where the assumption of normality was not fully met. Linearity of all self-report measures was confirmed via bivariate scatter plot analysis. Evaluation of the assumptions of homoscedasticity and non-multicollinearity/singularity were conducted prior to running each analysis containing these assumptions. These evaluations are mentioned and discussed in subsequent reporting in this section only when the assumptions were violated.

An alpha level of p = .05 was set as the significance criteria for all study analyses. A Bonferroni correction was not applied across multiple comparisons (e.g., follow-up tests to significant omnibus effects and interactions), as the nature of repeated measures within the study design rendered the alpha level impractically conservative when using the correction. That is, the Bonferroni correction for the coherence preference assessment blocks (8 repeated measures, $\alpha = .006$) and F-VAS administrations (17 repeated measures, $\alpha = .003$) was deemed to present an unacceptable risk of committing a type II error. Several contemporary critiques of experiment-wise alpha adjustments lend support to this analytic strategy (Nakagawa, 2004; Perneger, 1998; see also the discussion of alpha adjustments in Wilkinson & the APA Task Force on Statistical Inference, 1999).

In keeping with the recommendation of Cohen (1994; see also, Wilkinson & the APA Task Force on Statistical Inference, 1999), effect sizes were reported for all statistically significant effects. Partial eta squared (partial η^2) was reported for all mixed model ANOVAs, as it is the most appropriate estimate of the total variance accounted for by each predictor in a multifactor design (Pierce, Block, and Aguinois, 2004). Partial eta squared was also used in all other ANOVA models for the sake of consistency in effect size reporting. Cohen's *d* was reported for all significant comparisons of mean differences and Pearson's *r* was reported for all significant regression models. Phi (ϕ) was reported for all significant 2 by 2 chi-squared contingency tables and Cramer's V (ϕ_c) was reported for all significant 2 by 3 chi-squared contingency tables.

Participants

The frequency distribution of participant gender, race/ethnicity, and year in school for both the overall sample and each experimental condition is presented in Table 7. Chi-squared

tests of independence were used to evaluate differences in these characteristics between

experimental conditions. There was not a significant difference between conditions for gender,

$\chi^2(1, N = 81) = .001$, p = .980, $\phi = .003$. Fisher's exact test was used to evaluate differences

between groups for race/ethnicity and year in school as both of these participant characteristics

yielded expected cell counts of less than 5 in more than 20% of cells during chi-squared analysis.

There were no significant differences between conditions for race/ethnicity (p = .792) or year in

school (p = .138).

Table 7

Gender, Race/Ethnicity, and Year in School for Overall Sample and by Experimental Condition

Characteristic	Overall (n = 81) n (percent)	Meaning (n =44) n (percent)	Shapes (n = 37) n (percent)
Gender			
Male	22 (27.2%)	12 (27.3%)	10 (27.0%)
Female	59 (72.8%)	32 (72.7%)	27 (73.0%)
Race/Ethnicity			
Caucasian	56 (69.1%)	32 (72.7%)	24 (64.9%)
African American	21 (25.9%)	10 (22.7%)	11 (29.7%)
Asian	4 (4.9%)	2 (4.5%)	2 (5.4%)
Year in School			
Freshman	49 (60.5%)	23 (52.3%)	26 (70.3%)
Sophomore	17 (21.0%)	10 (22.7%)	7 (18.9%)
Junior	10 (12.3%)	6 (13.6%)	4 (10.8%)
Senior	5 (6.2%)	5 (11.4%)	0 (0.0%)

Mean values for age, initial self-report measures (i.e., GHQ-12, AAQ-II, CFQ, and

MCSDS-SF), practice test performance, and initial frustration level for both the overall sample

and each experimental condition are presented in Table 8. Two-tailed independent sample t-tests

were used to assess differences between experimental conditions on these variables (see Table 8

for t and p values from these analyses). There was a statistically significant difference between

conditions for social desirability (MCSDS-SF), t(79) = -2.36, p = .02, $d = 0.50$, with participants

in the Shapes condition displaying greater endorsement of socially desirable responses (M = 7.34,

SD = 2.93) than participants in the Meaning condition (M = 5.82, SD = 3.19). There were no

other significant differences between conditions on these study variables.

Table 8

Mean Scores on Initial Measures for Overall Sample and Comparison by Experimental Condition

Variable	Overall (n = 81)		Meaning (n = 44)		Shapes (n = 37)		t(79)	p
	M	SD	M	SD	M	SD		
Age	19.62	3.93	19.52	1.29	19.73	5.69	-0.24	.82
GHQ-12	1.69	2.59	1.57	2.64	1.84	2.56	-0.47	.64
AAQ-II	14.52	7.03	15.39	7.28	13.49	6.67	1.22	.23
CFQ	40.27	9.23	39.95	9.55	40.65	8.96	-0.03	.74
MCSDS-SF	6.56	3.16	5.82	3.19	7.43	2.93	-2.36	.02*
Practice Test	98.81	4.36	99.36	2.46	98.16	5.84	1.17[a]	.25
F-VAS 1	19.22	25.96	16.73	22.41	22.19	29.68	-0.94	.35

[a]Welch's t-test (df = 46.66) was used for this comparison as Levene's test for equality of variances was significant, indicating a departure from the assumption of homogeneity of variance between the two conditions.
* p < .05

Coherence Testing

A between subjects analysis was used in this phase to evaluate the effect of different

learning histories (i.e., Meaning and Shapes conditions) on performance during first and second

testing blocks. Given the intentionally ambiguous nature of the testing block, a response pattern

analysis was used to identify whether or not participants responded in patterns identified by the

researcher as coherent. In particular, participant responses for both testing blocks were scored

using both a shape based and meaning based scoring rubric. The number of responses scored as correct using each rubric served as the primary dependent measures. For the shape-scoring rubric, responses to each of the 27 Foods class stimuli were scored based on the topographical correspondence between the Foods class stimuli and the Shapes class stimuli. That is, matching line shaped food to the lines Shapes class member, circle shaped food to the circles Shapes class member, and triangle shaped food to the open triangles Shapes class member were scored as a correct response. Any other responding was scored as incorrect.

The meaning-scoring rubric was designed to determine whether or not participants assigned to the Meaning condition were spontaneously displaying transformation of stimulus functions. That is, the rubric detected whether participants treated the Shapes class members as if they have the functional properties of the Meaning class members. For example, the meaning scoring-rubric was sensitive to participant derived relations such as "the lines mean healthy, circles mean unhealthy, and the triangles must mean disgusting." There were six possible ways that participants could derive relations between Meaning class members and Shapes class members. The meaning-scoring rubrics scored each of the six ways and then extracted the combination that yielded the highest number of correct responses. Simulations of this rubric prior to participant contact demonstrated that if a participant perfectly followed a meaning based response strategy, one of the six meaning scores would be 100% with the other five varying between the value of 0% and 33%. Thus, retaining only the highest of the six meaning scores obtained was a valid and non-biased data reduction strategy.

Both a response pattern frequency analysis and a direct statistical comparison of scores generated by the meaning and shape scoring rubrics were used to determine if there were any

significant differences in response allocations between participants in the Shapes and Meaning conditions.

Response pattern analysis. For the response pattern frequency analysis, each participant was assigned to one of the following categories based on their responding during the testing block: meaning consistent, shape consistent, or other. If a participant obtained a score of 80% correct or above on the meaning scoring rubric their response pattern was classified as meaning consistent. If a participant obtained a score of 80% or above on the shape scoring rubric their response pattern was classified as shape consistent. If a participant obtained a score of 78.9% or lower on both the meaning and shape scoring-rubrics their response pattern was classified as other. Simulations of this classification system prior to participant contact demonstrated that a participant could only be sorted into one category. That is, it was impossible for a participant to score above an 80% on both the shape and meaning scoring rubrics.

The frequency distribution of response classifications during the first testing block is presented in Figure 9. Overall, 48 participants (59.26%) displayed either meaning or shape consistent responding in the first testing block and 33 participants (40.74%) displayed response patterns classified as other. A chi-square test of independence of response classifications during the first testing block yielded a statistically significant difference between conditions χ^2(2, N = 81) = 28.43, p < .001, ϕ_c = .592. A series of orthogonal follow up analyses were conducted to identify the source of the significant omnibus effect. There was a statistically significant difference between conditions in responses classified as other and responses classified as either meaning or shape consistent, χ^2(1, N = 81) = 16.4, p < .001, ϕ = .450. A greater proportion of Meaning condition participants displayed shape or meaning consistent responding (n = 35, 79.5%) compared to Shapes condition participants (n = 13, 35.1%). There was also a significant

difference between conditions in responses classified as meaning consistent and responses classified as shape consistent, $\chi^2(1, N = 48) = 15.1$, $p < .001$, $\phi = .561$. Among participants who displayed either meaning or shape consistent responding, those in the Shapes condition only displayed shape consistent responding (n =13, 100%) while those in the Meaning condition displayed both meaning consistent (n = 22, 62.9%) and shape consistent responding (n =13, 37.1%).
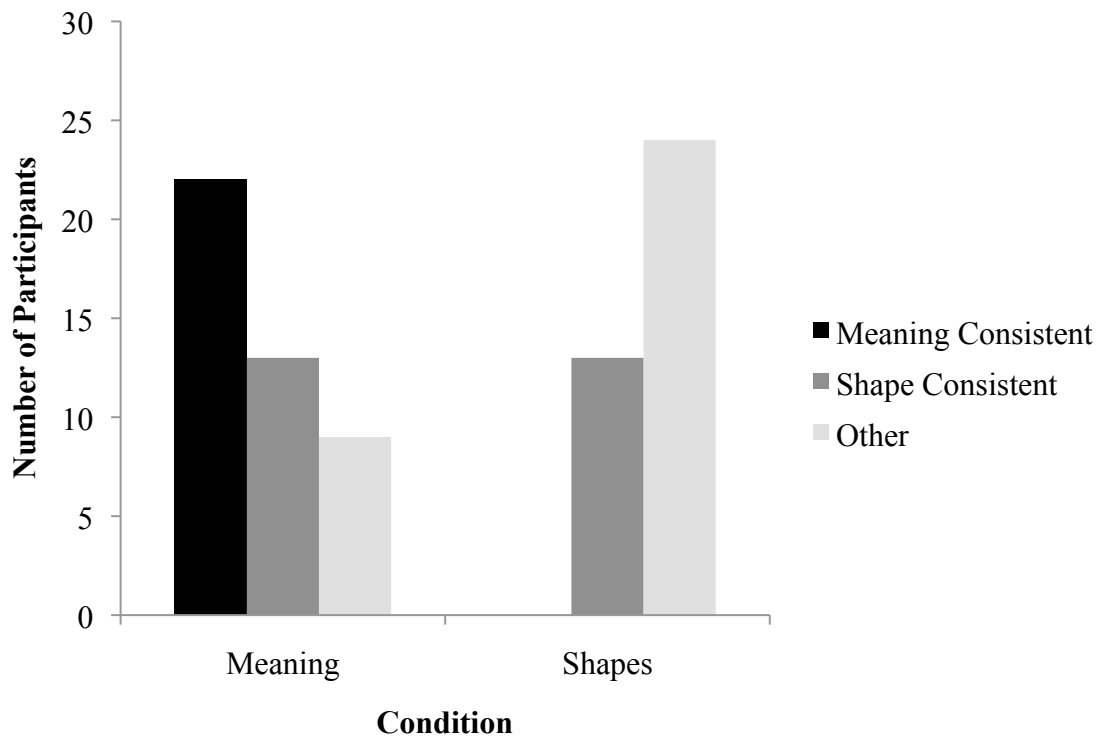
Figure 9. Frequency distribution of response classifications by experimental condition for the first testing block.

The frequency distribution of response classifications during the second testing block is presented in Figure 10. Overall, 59 participants (72.84%) displayed either meaning or shape consistent responding in the second testing block and 22 participants (27.16%) displayed response patterns classified as other. A chi-square test of independence of response classifications during the second testing block yielded a statistically significant difference

between conditions $\chi^2(2, N = 81) = 27.29$, $p < .001$, $\phi_c = .580$. The same orthogonal follow up

analysis used for the first testing block was applied to the second testing block. In the second

testing block there was a statistically significance difference between conditions in responses

classified as other and responses classified as either meaning or shape consistent, $\chi^2(1, N = 81) =$

6.16, $p = .013$, $\phi = .276$. A greater proportion of Meaning condition participants displayed

shape or meaning consistent responding ($n = 37$, 84.1%) compared to Shapes condition

participants ($n = 22$, 59.5%). There was also a significant difference between conditions in

responses classified as meaning consistent and responses classified as shape consistent, $\chi^2(1, N =$

59) = 22.4, $p < .001$, $\phi = .616$. Among participants who displayed either meaning or shape

consistent responding during the second testing block, those in the Shapes condition only

displayed shape consistent responding ($n = 22$, 100%) while those in the Meaning condition

displayed both meaning consistent ($n = 23$, 62.2%) and shape consistent responding ($n = 14$,

37.8%).
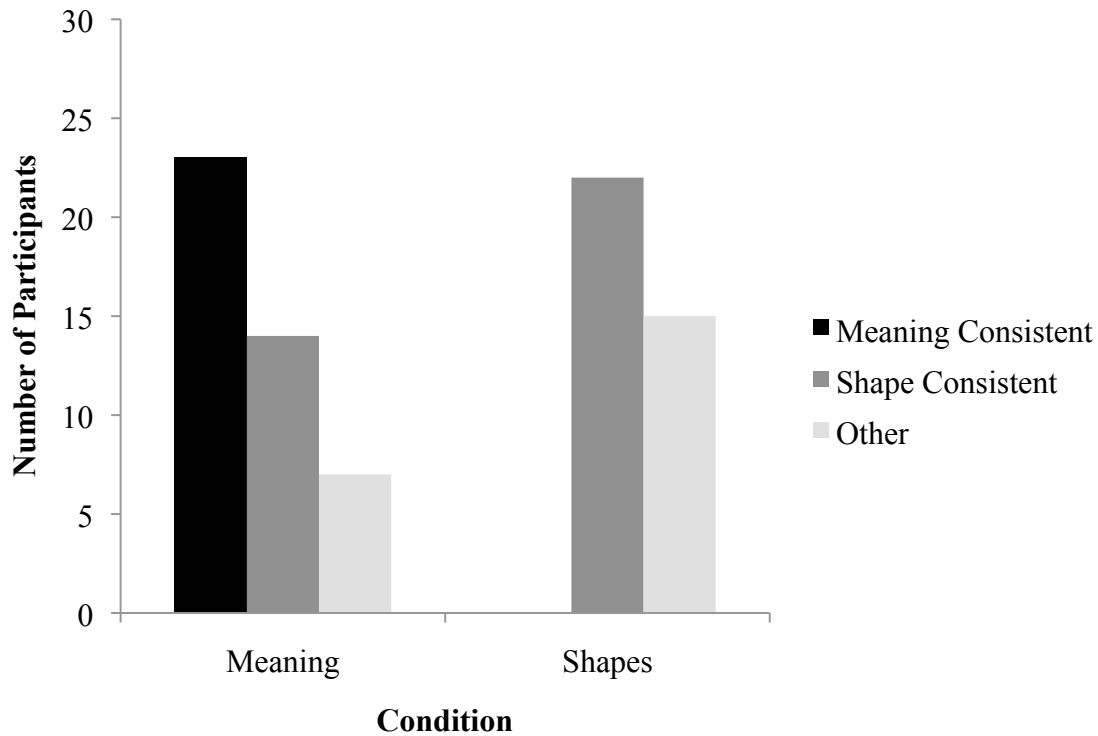
Figure 10. Frequency distribution of response classifications by experimental condition for the second testing block.

Direct statistical comparison. A direct comparison of percentage correct scores generated by the meaning and shape scoring rubrics was also conducted to assess differences between experimental conditions across both testing blocks. Mann-Whitney *U* tests were employed as the parametric assumption of homogeneity of variance was violated in these analyses due to the nature of how the scoring rubrics treated inaccurate responding. In the first testing block, a statistically significant difference between experimental conditions was found for scores generated by both the shape-scoring rubric, $U = 1,151.5$, p = .001, Z = 3.21, r = .357, and scores generated by the meaning-scoring rubric, $U = 413.5$, p < .001, Z = - 3.84, r = .416. Shapes condition participants scored higher (Mdn = 77.78) on the shape scoring rubric compared to meaning condition participants (Mdn = 38.89) and lower (Mdn = 40.74) on the meaning scoring rubric compared to meaning condition participants (Mdn = 74.07). A similar statistically

58

significant effect was found during the second testing block for both the shape-scoring rubric, U = 1,168, p = .001, Z = 3.37, r = .374, and meaning-scoring rubric, U = 405, p < .001, Z = - 3.91, r = .434. Within the second testing block, Shapes condition participants scored higher (Mdn = 85.19) on the shape scoring rubric compared to meaning condition participants (Mdn = 37.04) and lower (Mdn = 44.44) on the meaning scoring rubric compared to meaning condition participants (Mdn = 88.89).

Rule following analysis. The self-generated rules emitted by participants during the self-report task were coded into one of three categories: meaning based, shape based, or other. To be coded as meaning based, each participant's self-reported rules were required to contain a direct reference or close synonym to all three meaning based functions (i.e., healthy, unhealthy, and disgusting) and no reference to sorting based on other functions of the Foods class members. To be coded as shape based, each participant's self-reported rules were required to contain a direct reference or close synonym to all three shape topographies (i.e., line, circle, and triangles) and no reference to sorting based on other functions of the Foods class members. Self reported rules that contained a direct or close synonym to two shape topographies (e.g., line and circle) and either a direct negative definition (e.g. "triangle means not circles or lines") or an implied negative definition (e.g., triangle means disorganized/messy/unclear/confusing) for sorting Foods class members were also coded as shape based as long as they also made no reference to sorting based on other functions of the Foods class members. Self-reported rules that acknowledged an initial random or alternative response pattern but then clearly stated either a meaning based or shape based pattern were coded as meaning or shape based, respectively. Responses that did not meet criteria of the meaning based or shape based categories were categorized as other.

The primary investigator and a trained graduate student familiar with the project served as the primary and reliability coder, respectively. Both raters independently coded the self-reported rules and were blind to experimental condition and participant response patterns during the coding process. A reliability analysis using Cohen's Kappa statistic was conducted to determine the degree of consistency between raters. The inter-rater reliability was $\kappa = .716$, p <.001, 95% CI [.587, .845], indicating a substantial level of agreement following the guidelines established by Landis and Koch (1977).

The frequency distribution of self-reported rule classification by condition is presented in Figure 11. A chi-square test of independence yielded a statistically significant difference between conditions $\chi^2(2, N = 81) = 27.96$, p < .001, $\phi_c = .588$. A series of orthogonal follow up analyses were conducted to identify the source of the significant omnibus effect. There was not a significant difference between conditions when comparing rules classified as other to rules classified as either meaning or shape based, $\chi^2(1, N = 81) = .930$, p = .335, $\phi = .107$. This indicates that participants in both conditions did not differ in their generation and self-report of rules consistent with experimenter anticipated response strategies (i.e., meaning or shape based). However, there was a significant difference between conditions in rules classified as meaning based and rules classified as shape based $\chi^2(1, N = 61) = 27.4$, p < .001, $\phi = .670$. Of participants who generated coherent rules, those in the shape condition only generated shape based rules (n = 26, 100%) while those in the Meaning condition generated both meaning based (n = 23, 65.7%) and shaped based (n = 12, 34.3%) rules.
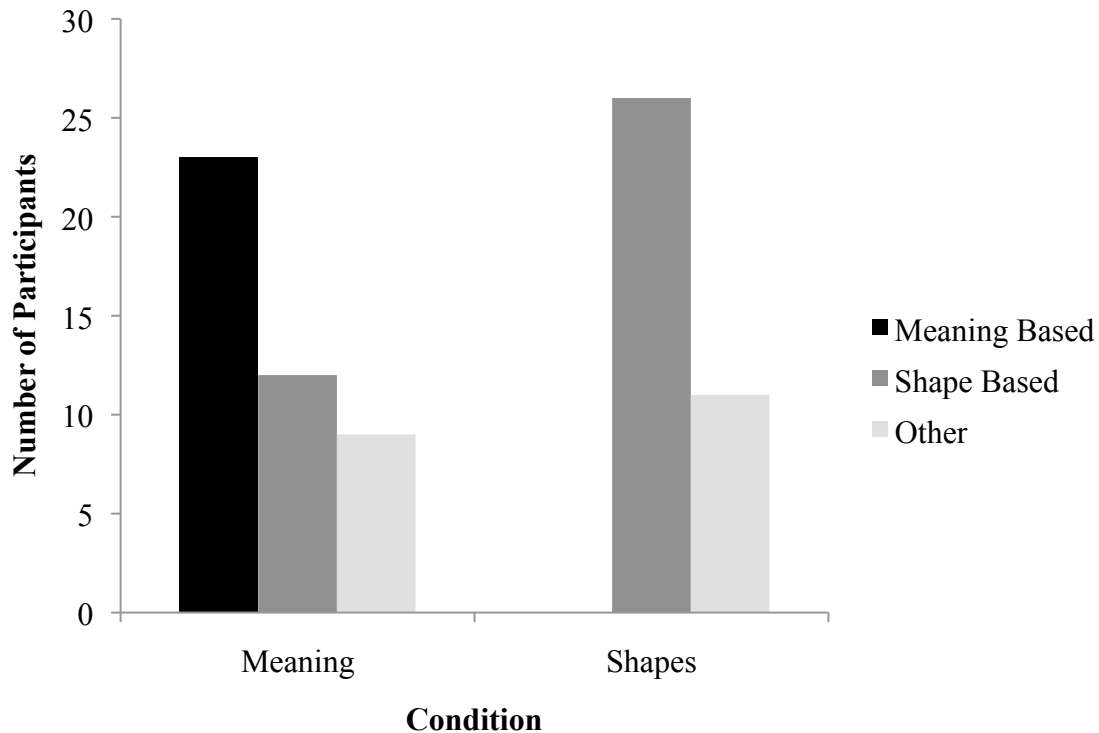
Figure 11. Frequency distribution of self-reported rule classifications by experimental condition.

A cross tabulation analysis was conducted to assess whether or not participants' self-generated rules were consistent with their previous responding. In this analysis, the consistency between response patterns observed during the first testing block (i.e., meaning consistent, shape consistent, and other) and the classification of self-generated rules (i.e., meaning based, shape based, or other) was assessed using Cohen's Kappa statistic. The cross tabulation matrix is presented in Table 9. There was moderate to substantial agreement between responding on the first testing block and self-generated rules, $\kappa = .614$, $p < .001$, 95% CI [.473, .755], with the majority of participants emitting self-generated rules that were consistent with their responding in the first testing block (n = 60, 74.07%). The most common inconsistency between response classification and self-generated rules was for participants who emitted rules classified as shape based but displayed response patterns categorized as other during the first testing block (n = 16, 19.75%). Other observed inconsistencies included rules classified as other but responses

classified as shape consistent (n = 4, 4.94%) and rules classified as meaning based but responses classified as other (n = 1, 1.23%).

Table 9.

Cross-tabulation of First Testing Block Response Patterns and Classifications of Self-Generated Rules (All Participants)

		Rule Classification			
		Meaning Based	Shape Based	Other	Total
Response Patterns	Meaning Consistent	22	0	0	22
	Shape Consistent	0	22	4	26
	Other	1	16	16	33
Total		23	38	20	81

A planned follow up analysis was then conducted to determine if the consistency between responding and self-generated rules differed as a function of experimental condition. Participants assigned to the Meaning condition (N = 44) displayed substantial agreement between their self-generated rules and their responding in the first testing block, $\kappa = .779$, $p < .001$, 95% CI [.622, .936], while participants assigned to the shapes condition (N = 37) displayed only fair agreement, $\kappa = .276$, $p = .031$, 95% CI [.051, .501]. The majority of Meaning condition participants emitted self-reported rules that were consistent with their response classification (n = 38, 86.36%,). However, three (6.82%) emitted other classified rules but displayed shape consistent responding, two (4.55%) emitted shape based rules but displayed other responding, and one (2.27%) emitted a meaning based rule but displayed other responding. While the majority of Shapes condition participants also emitted self-reported rules that were consistent with their responding in the first testing block (n = 22, 59.46%), there was a greater

proportion of inconsistencies for Shapes condition participants (n = 15, 40.54%) compared to

Meaning condition participants (n = 6, 13.64%). Of the 15 participants who displayed

inconsistencies in the Shapes condition, 14 (37.84%) emitted a shape based rule but an other

response pattern, and one (2.70%) emitted an other rule classification but a shape based response

pattern. The cross tabulation matrices for the meaning and shapes condition are presented in

Table 10 and Table 11, respectively.

Table 10.

Cross-tabulation of First Testing Block Response Patterns and Classifications of Self-Generated Rules (Meaning Condition)

		Rule Classification			
		Meaning Based	Shape Based	Other	Total
Response Patterns	Meaning Consistent	22	0	0	22
	Shape Consistent	0	10	3	13
	Other	1	2	6	9
Total		23	12	9	44

Table 11.

Cross-tabulation of First Testing Block Response Patterns and Classifications of Self-Generated Rules (Shapes Condition)

		Rule Classification			
		Meaning Based	Shape Based	Other	Total
Response Patterns	Meaning Consistent	0	0	0	0
	Shape Consistent	0	12	1	13
	Other	0	14	10	24
Total		0	26	11	37

To assess whether or not participants changed their responding to become more consistent with their self-generated rules, a Bhapkar test of marginal homogeneity was conducted. This analysis compared the response pattern classifications of the first testing block (administered before self-reporting of rules) to the response pattern classifications of the second testing block (administered after self-reporting of rules). The Bhapkar test is an alternative to the McNemar test of marginal homogeneity for contingency tables greater than 2 by 2 (Bhapkar, 1966). Tests of marginal homogeneity were used in this analysis instead of chi-squared tests, as response classifications across the two testing blocks were non-independent, thus violating a core assumption of the chi-squared test of independence. There was an overall significant effect between the testing blocks, Bhapkar $\chi^2(2, N = 81) = 9.05$, $p = .011$, indicating that there were significant changes in response classifications after the reporting of self-generated rules. There were no significant changes in meaning consistent classification across the two testing blocks (Fisher's exact $p = 1.0$). However, there were significant changes in both Shape consistent, McNemar $\chi^2 = 7.14$, $p = .008$, and other response classifications, McNemar $\chi^2 = 8.07$, $p = .005$. A cross tabulation analysis was then conducted to explore these significant changes in response classifications between the first and second testing block. The cross tabulation matrix is presented in Table 12. The majority of participants (n = 66, 81.48%) displayed consistent response classifications across the two testing blocks with 12 participants (14.81%) changing from other to shape, two participants (2.47%) changing from shape to other, and one participant (1.23%) changing from other to meaning based responding. There was substantial agreement between responding on the first testing block and responding on the second testing block, κ = .723, $p < .001$, 95% CI [.598, .848].

Table 12.

Cross-tabulation of First Testing Block Response Patterns and Second Testing Block Response Patterns (All Participants)

| | | Second Testing Block | | | |
		Meaning Consistent	Shape Consistent	Other	Total
First Testing Block	Meaning Consistent	22	0	0	22
	Shape Consistent	0	24	2	26
	Other	1	12	20	33
Total		23	36	22	81

Two planned follow up analyses were conducted to determine if the degree to which participants changed their responding across the two testing blocks differed as a function of experimental condition. There was not a significant difference in marginal homogeneity between blocks for the Meaning condition, Bhapkar χ^2(2, N = 44) = 1.38, p = .503, indicating that participants in the Meaning condition did not significantly alter their response patterns after self-reporting rules. However, there was a significant difference in marginal homogeneity between blocks for the Shapes condition, Bhapkar χ^2(2, N = 37) = 9.19, p = .010. Follow up analyses revealed significant changes in both shape consistent, McNemar χ^2 = 7.36, p = .007, and other response classifications, McNemar χ^2 = 7.36, p = .007, across the blocks. No Shapes condition participants showed meaning consistent responding in either testing block. In addition, a greater proportion of Shapes condition participants displayed shape consistent responding after the self-report task (n = 22, 59.5%) compared to before the self-report task (n = 13, 35.1%). See Table 13 and Table 14 for the cross tabulation matrices for the meaning and shapes condition, respectively.

Table 13.

Cross-tabulation of First Testing Block Response Patterns and Second Testing Block Response Patterns (Meaning Condition)

| | | Second Testing Block | | | |
		Meaning Consistent	Shape Consistent	Other	Total
First Testing Block	Meaning Consistent	22	0	0	22
	Shape Consistent	0	12	1	13
	Other	1	2	6	9
Total		23	14	7	44

Table 14.

Cross-tabulation of First Testing Block Response Patterns and Second Testing Block Response Patterns (Shapes Condition)

| | | Second Testing Block | | | |
		Meaning Consistent	Shape Consistent	Other	Total
First Testing Block	Meaning Consistent	0	0	0	0
	Shape Consistent	0	12	1	13
	Other	0	10	14	24
Total		0	22	15	37

Class Acquisition

Of the 81 participants who began the class acquisition matching to sample training, 77 (95.06%) successfully completed the training. Four participants (4.94%) failed to complete the training task and one (1.23%) withdrew from the study immediately after the training task. There was a statistically significant difference between completers and non-

66

completers/withdrawers on both total trial blocks to reach criterion, t(79) = 30.94, p < .001, $d =$ 12.52, and time in the training task t(79) = 15.65, p < .001, $d = 6.72$. Non-completers/withdrawers displayed a greater number of total trial blocks to reach criterion (M = 99, SD = 8.03) compared to completers (M = 9.43, SD = 6.16) and spent more time in the training task (M = 56.83, SD = 7.05) compared to completers (M = 12.63, SD = 6.06). For all subsequent study analyses, data from only the 76 completers was retained.

Participant performance during class acquisition is summarized in Table 15. Thirty participants (39.47%) failed to reach the 89% pass criterion on the test of combinatorial entailment while 46 participants (60.52%) emitted robust evidence of combinatorial entailment during the testing block. An exploratory stepwise regression analysis was conducted to determine if class acquisition training variables (i.e., trial blocks to criterion for the A-B, A-C, and Mixed A-B/A-C training phases, and time spent in training task) predicted performance on the test of combinatorial entailment. Total trial blocks to criterion were omitted from the analysis to prevent tolerance from being exceeded. Trial blocks to criterion for the initial A-B training phase was the only variable retained in the stepwise model and it significantly predicted combinatorial entailment performance, $\beta = -.319$, t = -2.90, p = .005. The overall model explained a significant amount of variance in combinatorial entailment scores, $R^2 = .102$, F(1,74) = 8.40, p = .005, with greater trial blocks to criterion in the A-B training phase associated with lower scores on the test of combinatorial entailment.

Table 15

Descriptive Statistics of Class Acquisition Performance

Variable	Mean	SD	Median	Min	Max
Trial Blocks to Criterion					
Train A-B	5.79	5.64	3	1	34
Train A-C	2.28	1.27	2	1	8
Mixed Train A-B/A-C	1.37	1.03	1	1	8
Total	9.43	6.16	7	3	38
Training Time (minutes)	12.63	6.06	11.07	6.14	42.43
Testing Accuracy (% correct)					
Mutual Entailment	92.91	11.84	100	39	100
Combinatorial Entailment	79.62	25.22	89	6	100

Note. Trial blocks to criterion were calculated for each participant by summing the number of times they were sequenced through the training block before meeting the pass criterion (≥89%).

Coherence Preference Assessment & Coherence Preference Assessment with Response Cost

These two phases were combined together for the purposes of data analysis. Both a response pattern analysis and a repeated measures statistical model were used to analyze participants' degree of preference towards coherent responding. For both analytic strategies, participant responses were assessed across the two concurrent choice blocks during the coherence preference assessment phase and the six concurrent choice blocks during the coherence preference assessment with response cost phase. The primary focus of both analytic techniques was to detect participants' preference towards coherent contexts during each of eight concurrent choice blocks presented in the study.

Response pattern analysis. For the response pattern analysis, participant responding across the eight concurrent choice blocks were categorized into one of the following categories: coherent preference, incoherent preference, and no preference. An error in the computer program systematically exposed all participants to 31 concurrent choice trials during each block

of the coherence testing phase instead of the intended 30, and 16 concurrent choice trials during each block of the coherence preference assessment with response cost phase instead of the intended 15. Given that all participants were equally exposed to the additional trial in each assessment block, there was no threat to the validity of the design and subsequent analyses were simply adjusted to account for exposure to the additional trials.

Participants were assigned to either the coherent or incoherent preference category if their individual response allocations within a block yielded a statistically significant chi-squared test value. That is, the obtained frequency distribution of responses for each participant in each block was compared to an expected frequency distribution of equal allocation to the coherent and incoherent initial links using a chi-squared goodness of fit test. To be classified in the coherent preference category, participants were required to select the coherent initial link 21 times or greater during the 31 trial blocks and 12 times or greater during the 16 trial blocks. To be classified in the incoherent preference category, participants were required to select the incoherent initial link 21 times or greater during the 31 trial blocks and 12 times or greater during the 16 trial. Participants were assigned to the no preference category if they failed to meet inclusion criteria for either of the categories above.

The frequency distribution of response pattern classifications across all eight blocks for all participants (n = 76) is presented in Figure 12. A multinomial goodness of fit exact test was performed on each assessment block to determine whether or not observed response classifications significantly deviated from random responding. An exact test was used instead of a chi-squared test as expected value counts for coherence preference and incoherence preference were less than five across all assessment blocks. Expected values were calculated to test the null hypothesis of chance responding (i.e., that each participant had a 50% chance of choosing the

69

coherent initial link on each trial). The Bernoulli process was used to determine the probability of a random responder being classified into the coherent preference, no preference, or incoherent preference category on the 16 trial blocks (n.b., the 16 trial blocks were chosen as they yielded a null hypothesis that was slightly more difficult to reject than the expected probabilities calculated from the 31 trial blocks). The resulting probabilities of 3.84% chance for coherent preference classification, 92.32% chance for no preference classification, and 3.84% chance for incoherent preference classification were used to weight expected values in the exact test model. Probability values of the multinomial goodness of fit exact test for each assessment block were all less than .001, resulting in a statistically significant departure from the null hypothesis. Thus, the null hypothesis of random responding was rejected for all eight assessment blocks, indicating that participant response classifications significantly deviated from those expected to be obtained under chance responding throughout the coherence preference assessment.
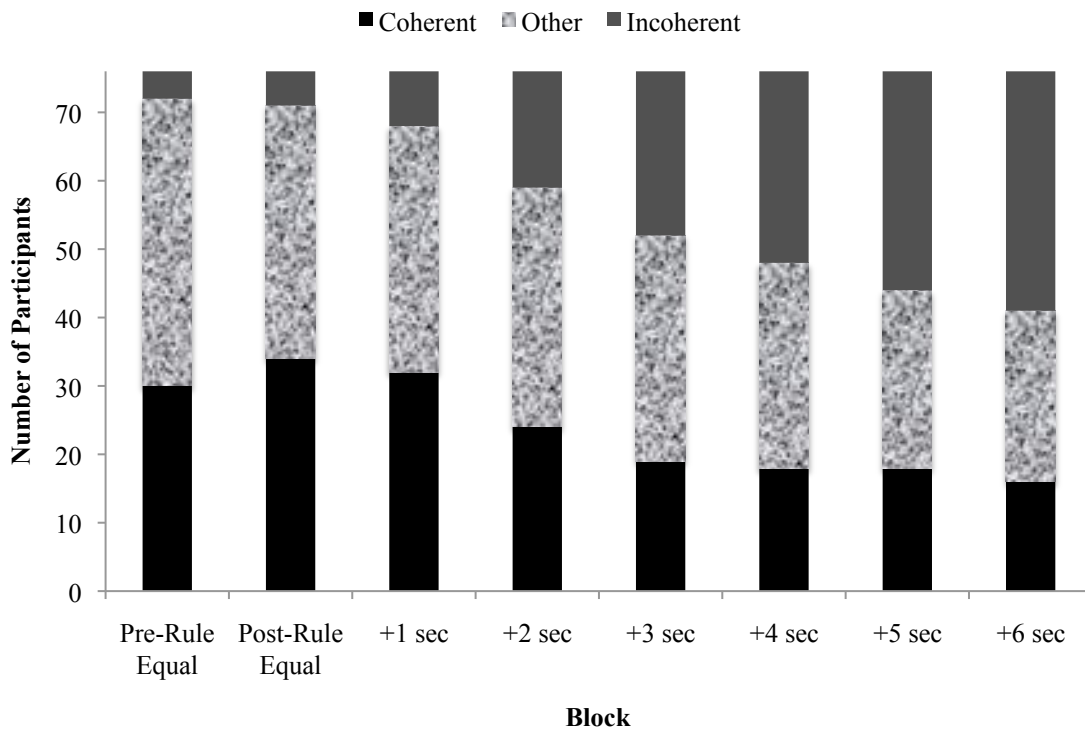
Figure 12. Frequency distribution of response pattern classifications across the eight coherent preference blocks for all participants (n = 76).

A follow up analysis was then undertaken to determine if there were differences between coherent and incoherent preference classification among participants who displayed a response preference (i.e., coherent or incoherence preference) within each assessment block. A Pearson's chi-squared goodness of fit test was used to assess departure from an expected frequency distribution of equal proportion coherent and incoherent preference classifications. Results for these analyses are presented in Table 16 along with frequency counts for coherent and incoherent classification. There was a significant departure from the null assumption of proportional response classification in the pre-rule equal, $\chi^2(1, N = 34) = 19.882$, p < .001, post-rule equal, $\chi^2(1, N = 39) = 21.564$, p < .001, and +1 second response cost assessment block, $\chi^2(1, N = 40) = 14.4$, p < .001. For all three blocks, there was a greater proportion of responses classified as

coherent; 88.24%, 87.18%, and 80% respectively. Response classifications in the +5 second

response cost and +6 second response cost assessment blocks were also statistically significant

from the null assumption of equal allocation, $\chi^2(1, N = 50) = 3.920$, p = .048, and $\chi^2(1, N = 51)$

= 7.078, p .008, respectively. Response classification trended in the opposite direction than in

earlier blocks, with 64% and 68.63% percent of responses classified as incoherent, respectively.

In addition, there was an overall linear trend of increased percentage of participants displaying

either coherent or incoherent response classifications across the eight assessment blocks with

67.11% of the total sample displaying a preference in the +6 second response cost block

compared to only 44.74% of the total sample in the pre-rule equal assessment block.

Table 16

*Comparison of Frequency Counts of Coherent and Incoherent Classifications across the Eight
Assessment Blocks for All Participants (n = 76)*

Assessment Block	Response Classification		n Displaying Preference (% of Total N)	$\chi^2(1)$	p
	Coherent n (%)	Incoherent n (%)			
No Response Cost					
Pre-Rule Equal	30 (88.24%)	4 (11.76%)	34 (44.74%)	19.882	<.001*
Post-Rule Equal	34(87.18%)	5 (12.82%)	39 (51.32%)	21.564	<.001*
Response Cost					
+ 1 second	32 (80.00%)	8 (20.00%)	40 (52.63%)	14.400	<.001*
+ 2 second	24 (58.54%)	17 (41.46%)	41 (53.95%)	1.195	.274
+ 3 second	19 (44.19%)	24 (55.81%)	43 (56.58%)	0.581	.446
+ 4 second	18 (39.13%)	28 (60.87%)	46 (60.53%)	2.174	.140
+ 5 second	18 (36.00%)	32 (64.00%)	50 (65.80%)	3.920	.048*
+ 6 second	16 (31.37%)	35 (68.63%)	51 (67.11%)	7.078	.008*

* p < .05

Response pattern analysis moderated by combinatorial entailment performance. In

addition to the full sample response pattern analysis conducted above, two follow up response

pattern analyses were conducted to explore response classifications between participants who

passed (n = 46) and who failed (n = 30) the combinatorial entailment test in the class acquisition

phase. Combinatorial entailment was expected to moderate response classifications, as the functional properties of the contextual cues (i.e., coherent and non-coherent) were theoretically assumed to only be fully salient for the participants who displayed robust combinatorial entailment.

Combinatorial entailers. The frequency distribution of response pattern classifications across all eight blocks for participants who passed the test of combinatorial entailment (n = 46) is presented in Figure 13. A multinomial goodness of fit exact test, identical to the one described in the response pattern analysis section above, was performed on each assessment block to determine whether or not observed response classifications significantly deviated from random responding for combinatorial entailers. Probability values of the multinomial goodness of fit exact test for each assessment block among combinatorial entailers were all less than .001, resulting in a statistically significant departure from the null hypothesis. Thus, the null hypothesis of random responding was rejected for all eight assessment blocks, indicating that the response classifications of combinatorial entailers significantly deviated from those expected to be obtained under chance responding throughout the coherence preference assessment.
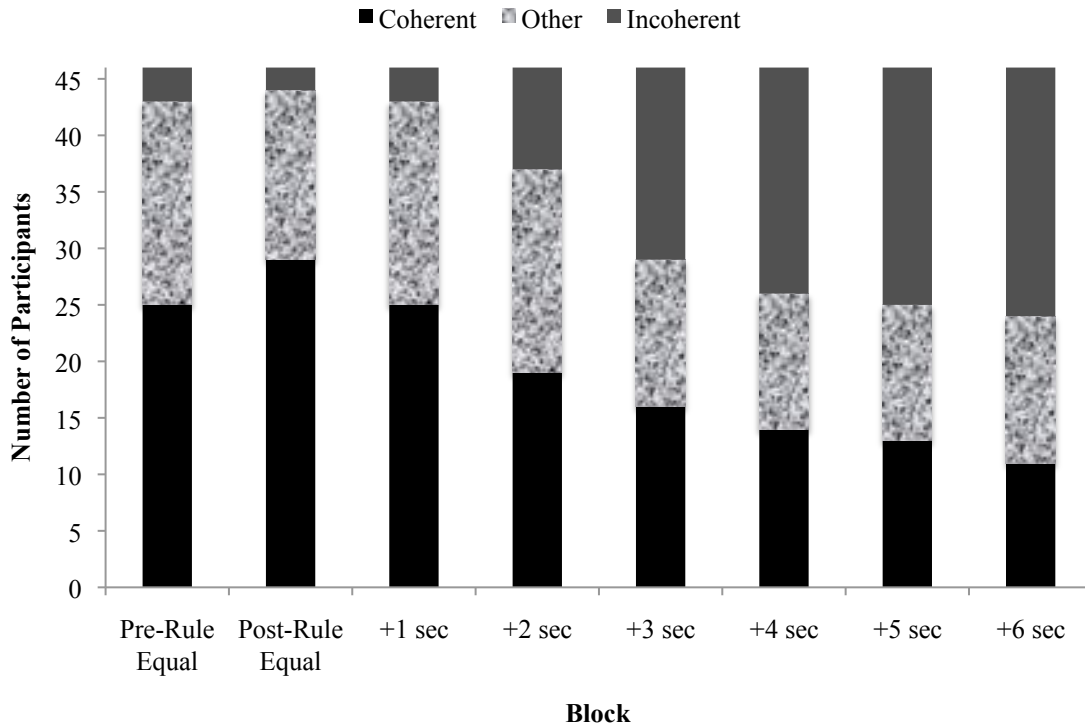
Figure 13. Frequency distribution of response pattern classifications across the eight coherent preference blocks for participants who passed the test of combinatorial entailment (n = 46).

A follow up analysis was then undertaken to determine if there were differences between coherent and incoherent preference classification among combinatorial entailers who displayed a response preference (i.e., coherent or incoherence preference) within each assessment block. A Pearson's chi-squared goodness of fit test was used to assess departure from an expected frequency distribution of equal proportion coherent and incoherent preference classifications. Results for these analyses are presented in Table 17 along with frequency counts for coherent and incoherent classification. There was a significant departure from the null assumption of proportional response classification for entailers in the pre-rule equal, $\chi^2(1, N = 28) = 17.286$, p < .001, post-rule equal, $\chi^2(1, N = 31) = 23.516$, p < .001, and +1 second response cost assessment block, $\chi^2(1, N = 28) = 17.286$, p < .001. For all three blocks, there was a greater proportion of responses classified as coherent: 89.29%, 93.55%, and 89.29% respectively. There

was also a reduction in the proportion of coherent response classifications and a concomitant increase in the proportion of incoherent response classifications across the eight assessment blocks. However, this trend did not result in significant differences between response classifications in later assessment blocks. A relatively stable proportion of entailers displayed either coherent or incoherent response classifications across the eight assessment blocks (range 60.87% to 73.92%).

Table 17

Comparison of Frequency Counts of Coherent and Incoherent Classifications across the Eight Assessment Blocks for Combinatorial Entailers (n = 46)

| Assessment Block | Response Classification | | n Displaying Preference (% of Total N) | χ^2 (1) | p |
	Coherent n (%)	Incoherent n (%)			
No Response Cost					
Pre-Rule Equal	25 (89.29%)	3 (10.71%)	28 (60.87%)	17.286	<.001*
Post-Rule Equal	29 (93.55%)	2 (6.45%)	31 (67.39%)	23.516	<.001*
Response Cost					
+ 1 second	25 (89.29%)	3 (10.71%)	28 (60.87%)	17.286	<.001*
+ 2 second	19 (67.86%)	9 (32.14%)	28 (60.87%)	3.571	.059
+ 3 second	16 (48.48%)	17 (51.51%)	33 (71.74%)	0.030	.862
+ 4 second	14 (41.18%)	20 (58.82%)	34 (73.92%)	1.059	.303
+ 5 second	13 (38.24%)	21 (61.76%)	34 (73.92%)	1.882	.170
+ 6 second	11 (33.33%)	22 (66.67%)	33 (71.74%)	3.667	.056

* p < .05

Non-combinatorial entailers. The frequency distribution of response pattern classifications across all eight blocks for participants who failed the test of combinatorial entailment (n = 30) is presented in Figure 14. A multinomial goodness of fit exact test identical to the one described for combinatorial entailers was performed on each assessment block to determine whether or not observed response classifications significantly deviated from random responding for combinatorial entailers. The probability value of the multinomial goodness of fit

exact test for the pre-rule equal assessment block was statistically significant, but at a much larger probability value (p = .014) than the ones obtained for all exact tests run for the combinatorial entailers (p < .0000001 for all eight assessment blocks). This finding indicated that, as a group, the response classifications of non-entailers was closer to random responding during the first assessment block (16.67% classified as coherent responders, 80% as other responders, and 3.33% as incoherent responders) than the classifications observed for entailers. In addition, the probability value of the exact test for non-entailers in the post-rule equal assessment block was significant at p = .001. During the post-rule equal assessment block 16.67% of non-entailers were classified as coherent responders, 73.33% as other responders, and 10% as incoherent responders. Fisher's exact tests for entailers across the remaining six assessment blocks were significant at p < .0001, indicating a significant departure from random responding across all the response cost preference assessment blocks.
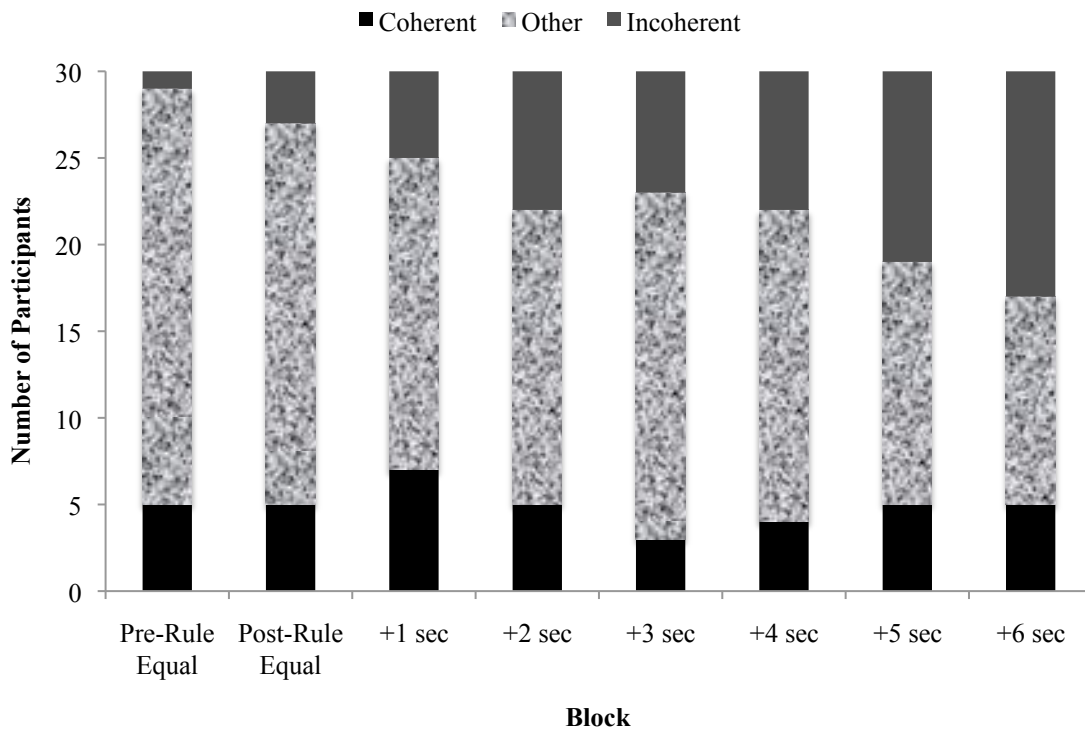
Figure 14. Frequency distribution of response pattern classifications across the eight coherent preference blocks for participants who failed the test of combinatorial entailment (n = 30).

A follow up analysis was then conducted to determine if there were significant differences between coherent and incoherent preference classification among non-entailers who displayed a response preference within each assessment block. A Pearson's chi-squared goodness of fit test was used to assess departure from an expected frequency distribution of equal proportion coherent and incoherent preference classifications. Results for these analyses are presented in Table 18 along with frequency counts for coherent and incoherent classification. No significant departures between the obtained classifications and the null assumption of equal allocation of classifications were found. There was a general linear trend of increased proportion of response classified as incoherent across the eight assessment blocks for non-entailers with 72.22% classified as incoherent in the + 6 second response cost block compared to only 16.67% classified as incoherent in the pre-rule equal assessment block. There was also a linear trend for

percentage of non-entailers displaying either coherent or incoherent response classifications across the eight assessment blocks with 60% of non-entailers displaying a preference in the +6 second response cost block compared to only 20% of the entailers in the Pre-Rule equal assessment block.

Table 18

Comparison of Frequency Counts of Coherent and Incoherence Classifications across the Eight Assessment Blocks for Non-Combinatorial Entailers (n = 30)

| Assessment Block | Response Classification | | n Displaying Preference (% of Total N) | $\chi^2(1)$ | p |
	Coherent n (%)	Incoherent n (%)			
No Response Cost					
Pre-Rule Equal	5 (83.33%)	1 (16.67%)	6 (20.00%)	exact[a]	.213
Post-Rule Equal	5 (62.5%)	3 (37.5%)	8 (26.67%)	exact[a]	.726
Response Cost					
+ 1 second	7 (58.33%)	5 (41.67%)	12 (40.00%)	0.333	.564
+ 2 second	5 (38.46%)	8 (61.54%)	13 (43.33%)	0.692	.405
+ 3 second	3 (30.00%)	7 (70.00%)	10 (33.33%)	1.600	.206
+ 4 second	4 (33.33%)	8 (66.67%)	12 (40.00%)	1.333	.248
+ 5 second	5 (31.25%)	11(68.75%)	16 (53.33%)	2.250	.134
+ 6 second	5 (27.78%)	13 (72.22%)	18 (60.00%)	3.556	.059

[a] Fisher's exact test was used for this comparison as the expected count was less than 5 for both classifications.
* $p < .05$

Direct comparison of combinatorial entailers and non-combinatorial entailers. To determine whether response preference classifications of participants differed as a function of performance on the test of combinatorial entailment, a series of chi-squared tests of independence (N = 76) were conducted across the eight preference assessment blocks. In each block, the frequency distribution of response classifications (coherent, other, and incoherent) for entailers and non-entailers were compared. See Table 19 for the obtained results and frequency counts. A significant omnibus effect for entailers compared to non-entailers was found for the

78

pre-rule equal assessment block, post-rule equal assessment block, + 1 second response cost

block, +3 second response cost block, and +4 second response cost block.

Table 19

Comparison of Response Classifications across the Eight Preference Assessment Blocks by Combinatorial Entailment Status

Assessment Block	Response Classification			$\chi^2 (2)$	p	ϕ_c
	Coherent n (%)	Other n (%)	Incoherent n (%)			
Pre-Rule Equal				exact[a]	<.001*	-
Entailers	25 (54.35%)	18 (39.13%)	3 (6.52%)			
Non-Entailers	5 (16.67%)	24 (80.00%)	1 (3.33%)			
Post-Rule Equal				exact[a]	<.001*	-
Entailers	29 (63.04%)	15 (32.61%)	2 (4.35%)			
Non-Entailers	5 (16.67%)	22 (73.33%)	3 (10.00%)			
+ 1 second				exact[a]	.021*	-
Entailers	25 (54.35%)	18 (39.13%)	3 (6.52%)			
Non-Entailers	7 (23.33%)	18 (60.00%)	5 (16.67%)			
+ 2 second				5.11	.078	.259
Entailers	19 (41.30%)	18 (39.13%)	9 (19.57%)			
Non-Entailers	5 (16.67%)	17 (56.67%)	8 (26.67%)			
+ 3 second				11.7	.003*	.392
Entailers	16 (34.78%)	13 (28.26%)	17 (36.96%)			
Non-Entailers	3 (10.00%)	20 (66.67%)	7 (23.33%)			
+ 4 second				8.93	.012*	.343
Entailers	14 (30.43%)	12 (26.09%)	20 (43.48%)			
Non-Entailers	4 (13.33%)	18 (60.00%)	8 (26.67%)			
+ 5 second				3.63	.163	.219
Entailers	13 (28.26%)	12 (26.09%)	21 (45.65%)			
Non-Entailers	5 (16.67%)	14 (46.67%)	11 (36.67%)			
+ 6 second				1.29	.524	.130
Entailers	11 (23.91%)	13 (28.26%)	23 (50.00%)			
Non-Entailers	5 (16.67%)	12 (40.00%)	13 (43.33%)			

[a]Fisher's exact probability test was used for these comparisons as greater than 20% of cells had expected values of less than 5.
* p < .05

A series of orthogonal follow up analyses were conducted to identify the source of each

significant omnibus effect. For each significant omnibus effect, a chi-squared test of

independence comparing response patterns classified as other to response patterns classified as

either coherent or incoherent by combinatorial entailment status was conducted. In addition, a chi-squared test of independence comparing responses patterns classified as coherent to response patterns classified as incoherent by combinatorial entailment status was also conducted. Fisher's exact probability tests were substituted for the planned comparisons whenever greater than 20% of cells had expected values of less than 5.

In the pre-rule equal assessment block, there was a significant difference between response patterns classified as other and response patterns classified as either coherent or incoherent preference by combinatorial status, $\chi^2(1, N = 76) = 12.27$, $p < .001$, $\phi = .402$. Participants who failed the test of combinatorial entailment displayed a greater proportion of other response patterns (66.67%) compared to participants who passed the test of combinatorial entailment (28.26%). The response patterns of entailers and non-entailers who displayed a response preference did not differ between coherent and incoherent preference ($p = .559$).

In the post-rule equal assessment block, there was a significant difference between response patterns classified as other and response patterns classified as either coherent or incoherent preference by combinatorial status, $\chi^2(1, N = 76) = 12.05$, $p < .001$, $\phi = .398$. Participants who failed the test of combinatorial entailment displayed a greater proportion of other response patterns (73.33%) compared to participants who passed the test of combinatorial entailment (32.61%). The response patterns of entailers and non-entailers who displayed a response preference differed significantly ($p = .049$), with a greater proportion of entailers (93.55%) displaying a coherent preference compared to non-entailers (62.5%)

In the +1 second response cost assessment block, there was not a significant difference between response patterns classified as other and response patterns classified as either coherent or incoherent preference by combinatorial status, $\chi^2(1, N = 76) = 3.17$, $p = .075$, $\phi = .204$.

However, the response patterns of entailers and non-entailers who displayed a response preference did differ between coherent and incoherent preference (p = .039), with entailers (89.29%) displaying a greater proportion of coherent preference compared to non-entailers (58.33%).

In the +3 second response cost assessment block, there was a significant difference between response patterns classified as other and response patterns classified as either coherent or incoherent preference by combinatorial status, $\chi^2(1, N = 76) = 10.9$, p = .001, $\phi = .379$. Participants who failed the test of combinatorial entailment displayed a greater proportion of other response patterns (73.33%) compared to participants who passed the test of combinatorial entailment (32.61%). The response patterns of entailers and non-entailers who displayed a response preference did not differ significantly between coherent and incoherent preference (p = .470).

In the +4 second response cost assessment block, there was a significant difference between entailers and non-entailers in the frequency distribution of response patterns classified as other and response patterns classified as either coherent or incoherent preference, $\chi^2(1, N = 76) = 8.74$, p = .003, $\phi = .339$. Non-entailers displayed a greater proportion of other response patterns (60%) compared to entailers (26.09%). The response patterns of entailers and non-entailers who displayed a response preference did not differ significantly between coherent and incoherent preference (p = .739).

Repeated measures analysis. Prior to running this statistical model, the percentage of responses each participant allocates towards the coherent contextual cue was calculated for each of the eight concurrent choice blocks. The use of a percentage as the primary dependent measure allowed for meaningful comparisons of preference across all eight blocks despite the fact that

some blocks differ in the number of trials. An eight-time point repeated measures ANOVA was conducted to determine if there was an overall main effect for percentage allocated towards coherent contexts across the eight current choice blocks. Mauchly's test of sphericity was significant within the model, $W = .017$, $\chi^2(27) = 295.668$, $p < .001$, Greenhouse-Geisser $\varepsilon = .376$, indicating a departure from the assumption of equality of variances/covariance patterns in the variance/covariance matrix of the observed data. A Greenhouse-Geisser correction was applied to mitigate the violation of this assumption.

There was a significant main effect across the eight assessment blocks, $F(2.632, 197.348) = 18.768$, $p < .001$, partial $\eta^2 = .200$, indicating changes in percentage of responding allocated towards the coherent contextual cue across the blocks. Figure 15 presents mean percentage coherent values at each time point along with standard error. A follow up polynomial contrast analysis revealed a significant linear, $F(1, 75) = 32.087$, $p < .001$, partial $\eta^2 = .300$, and cubic, $F(1, 75) = 11.608$, $p = .001$, partial $\eta^2 = .134$, trend in the obtained data. These contrasts indicate that, on average, participants decreased their allocation towards the coherent contextual cue across the assessment blocks (linear trend) and slightly increased their allocation of responses towards coherent contextual cue between the pre-rule equal and post-rule equal before decreasing their allocation across the six response cost assessment blocks (cubic trend).

Figure 15. Percentage of responses allocated towards the coherent contextual cue across the eight preference assessment blocks.

A follow up analysis was conducted across the eight assessment blocks to determine whether or not the obtained allocation of responding towards the coherent contextual cue was significantly greater or less than the allocation of responding hypothesized to occur under random response allocation (i.e., a 50% chance of choosing the coherent contextual cue on each concurrent choice trial). A series of one-sample t-tests were conducted for each assessment block comparing the obtained response distribution to an identical distribution with a mean set at 50. The mean difference of the two distributions, inferential test results, and the estimated upper and lower bounds of the 95% confidence interval of the difference for each assessment block is presented in Table 20. Obtained findings indicated a significant departure from the null hypothesis of random responding during the pre-rule equal, post-rule equal, and +1 second response cost blocks with a greater proportion of responses allocated towards the coherent cue.

An opposite significant effect was observed during the +5 second and +6 second response cost

blocks with a lower proportion of responses allocated towards the coherent cue than would be

expected under random responding.

Table 20

Comparison of Obtained Distributions of Responses Allocated Towards the Coherent Contextual Cue to a Null Model Assuming Random Responding

Assessment Block	Obtained Responses		Mean Diff.	t(76)	p	d	95% Confidence Interval of Difference	
	M	SD					LL	UL
No Response Cost								
Pre-Rule Equal	64.07	23.33	14.07	5.26	<.001*	0.60	8.73	19.40
Post-Rule Equal	68.12	26.57	18.12	5.95	<.001*	0.68	12.05	24.19
Response Cost								
+ 1 second	65.24	29.91	15.24	4.44	<.001*	0.51	8.40	22.07
+ 2 second	55.50	35.28	5.50	1.36	.178	0.16	-2.56	13.56
+ 3 second	49.57	35.22	-.43	-.107	.915	-0.01	-8.48	7.61
+ 4 second	44.11	36.01	-5.90	-1.43	.158	-0.16	-14.12	2.33
+ 5 second	40.41	37.44	-9.59	-2.23	.028*	-0.26	-18.15	-1.04
+ 6 second	38.58	36.83	-11.42	-2.70	.008*	-0.31	-19.84	-3.01

* p < .05

A series of planned moderation analyses were then conducted to explore the significant

main effect found for all participants. First, a manipulation check was performed to determine

whether or not exposure to the different experimental conditions in the coherence testing phase

of the study moderated allocation of coherent responding during the preference assessment

blocks. An eight (assessment blocks) by two (Meaning or Shapes condition) mixed ANOVA

model was used for this analysis. The assumption of sphericity was violated in this model, W

$= .017$, $\chi^2(27) = 293.237$, $p < .001$, Greenhouse-Geisser $\varepsilon = .375$, and a Greenhouse-Geisser

correction was applied to mitigate the violation. In addition, Box's test of equality of covariance

matrices was significant, $M = 74.871$, F(36, 16789.74) = 1.836, p = .002, indicating that the

assumption of homoscedasticity was not met. However, Box's M is extremely sensitive to minor

84

deviations from multivariate normality, especially in a design with unequal cell sizes (Tabachnick & Fidell, 2007). Given that the significance value of the test did not meet the $p < .001$ alpha criterion recommended by Tabachnick and Fidell (2007), the robustness of this model to the assumption of homoscedasticity was assumed. There was not a main effect for experimental condition on coherent response allocation, $F(1, 74) = .180$, $p = .673$, partial $\eta^2 = .002$, nor was there a significant interaction between experimental condition and assessment block, $F(2.627, 194.402) = .689$, $p = .541$, partial $\eta^2 = .009$. The main effect of assessment blocks remained significant in the model, $F(2.627, 194.402) = 19.041$, $p < .001$, partial $\eta^2 = .205$ These findings indicate that exposure to different experimental conditions during the coherence testing phase of the study did not affect participant performance during the coherence preference assessment blocks.

The potential moderator of social desirability on allocation of coherent responding was then explored using an eight time point (assessment blocks) repeated measures ANOVA with MCSDS-SF scores entered as a continuous covariate predictor. The assumption of sphericity was violated in this analysis, $W = .017$, $\chi^2(27) = 293.110$, $p < .001$, Greenhouse-Geisser $\varepsilon = .373$ and a Greenhouse-Geisser correction was applied to mitigate the violation. There was not a significant effect for the social desirability by assessment block interaction, $F(2.609, 193.072) = 1.552$, $p = .208$, partial $\eta^2 = .021$, and the main effect for assessment block remained significant in the model, $F(2.609, 193.072) = 7.276$, $p < .001$, partial $\eta^2 = .090$. There was a significant main effect for social desirability, $F(1, 74) = 5.381$, $p = .023$, partial $\eta^2 = .068$. A follow-up linear regression was conducted to determine the direction of the effect. MCSDS-SF scores were entered as the predictor variable and the sum of each participant's percentage coherent value across the eight assessment blocks were entered as the predicted variable. Greater levels of

social desirability were associated with greater overall coherent preference, $\beta = .260$, t = 2.320, p = .023.

Performance on the test of combinatorial entailment was assessed as a potential moderator of preference towards coherent contexts using an eight time point (assessment blocks) by two (entailers and non-entailers) mixed model ANOVA. While performance on the test of combinatorial entailment was continuous (i.e., percentage score from 0 to 100), a dichotomized measure of performance was utilized as a benchmark of 89% or greater accuracy is commonly used within matching to sample literature (Green & Saunders, 1998). The assumption of sphericity was violated in this model, $W = .019$, $\chi^2(27) = 283.350$, $p < .001$, Greenhouse-Geisser $\varepsilon = .389$, and a Greenhouse-Geisser correction was applied to mitigate the violation. In addition, Box's test of equality of covariance matrices was significant, $M = 80.460$, F(36, 13030.182) = 1.956, p = .001, indicating that the assumption of homoscedasticity was not met. However, the robustness of this model to the assumption of homoscedasticity was assumed as the significance value of the test did not meet the p < .001 alpha criterion recommended by Tabachnick and Fidell (2007).

There was a significant interaction of assessment block by combinatorial entailment performance, F(2.721, 201.367) = 3.083, p = .033, partial $\eta^2 = .040$, and the main effect of assessment block remained significant in the model, F(2.721, 201.367) = 15.604, p < .001 , partial $\eta^2 = .174$. In addition, the main effect for combinatorial entailment performance was approaching significance, F(1, 74) = 3.728, p = .057, partial $\eta^2 = .174$, with non-entailers on average displaying fewer responses allocated towards coherence (M = 46.61, SE = 4.39) compared to entailers (M = 57.49, SE = 3.54). Figure 16 presents mean percentage of responses

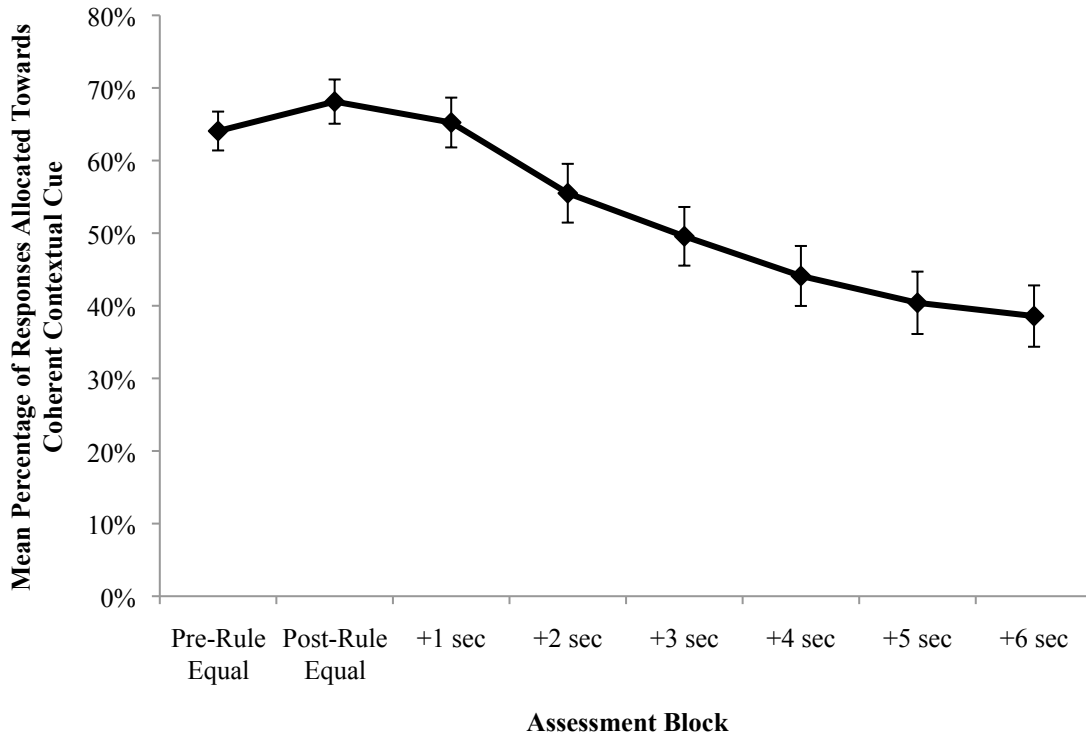allocated towards the coherent contextual cue at each assessment point by combinatorial

entailment status.

Figure 16. Percentage of responses allocated towards the coherent contextual cue across the eight preference assessment blocks by combinatorial entailment performance. Entailers (n = 46) passed the test of combinatorial entailment with a score of 89% or greater while Non-Entailers (n =30) failed the test of combinatorial entailment.

A follow up analysis was conducted to explore the significant interaction effect of

assessment block by combinatorial entailment performance. Independent sample t-tests were

used to compare entailer and non-entailer response allocations towards coherence across each of

the eight preference assessment blocks. There was a significant difference between entailers and

non-entailers during the pre-rule equal assessment block, t(73.92) = -4.158, p < .001, *d* = 0.93,

with entailers (M = 71.61, SD = 24.63) displaying a significantly greater response allocation

towards the coherent cue compared to non-entailers (M = 52.50, SD = 15.43). Similar

significant effects were also found in both the post-rule equal assessment block, t(73.08) = -

4.416, p < .001 , d = 1.0, and +1 second response cost block, t(74) = -2.997, p = .004, d = 0.71.

Entailers displayed a greater response allocation towards coherence in the post-rule equal block

(M = 77.35, SD = 26.72) compared to non-entailers (M = 53.97, SD = 19.38). Entailers also

displayed a greater response allocation towards coherence in the +1 second response cost block

(M = 73.13, SD = 29.52) compared to non-entailers (M = 53.13, SD = 26.65). No other

significant differences emerged between entailers and non-entailers indicating that while the two

groups differed in response allocations towards coherence during the early assessment blocks,

they did not differ significantly in later blocks. This observation is supported by the presence of

both a significant linear contrast of the assessment block by combinatorial entailment

performance interaction, F(1, 74) = 4.444, p = .038, partial η^2 = .057, and a significant cubic

contrast of the interaction, F(1, 74) = 4.001, p = .049, partial η^2 = .051.

Rule following analysis. The rules generated by participants during the self-report task

were coded into one of two categories: accurate or other. All responses that identified the

coherent contextual cue as leading to coherent contexts and the incoherent contextual cue as

leading to incoherent contexts were coded as accurate (e.g., self reports that referenced one color

being easier, less ambiguous, correct, valid, true, good, less frustrating, similar, etc.). Responses

were coded as accurate if they included any substantial reference to the distinction between

coherent and incoherent contexts, even if parts of the self-reported rule were unrelated.

Responses that did not meet criteria of the accurate category were categorized as other.

The primary investigator and a trained graduate student familiar with the scope of the

project served as the primary and reliability coder, respectively. Both raters independently coded

the self-reported rules and were blind to participant response patterns on the matching to sample

and concurrent chain task during the coding process. A reliability analysis using Cohen's Kappa

statistic was conducted to determine the degree of consistency between raters. The inter-rater reliability for the ratings was found to be κ = .967, p < .001, 95% CI [.90, 1.0], indicating almost perfect agreement following the guidelines established by Landis and Koch (1977).

A cross tabulation analysis comparing rule classification (accurate or other) to combinatorial entailment performance (entailer and non-entailer) was conducted to determine whether self-report of accurate rules required a display of combinatorial entailment. Of the 20 participants who emitted accurate rules, 17 (85%) were classified as entailers and three (15%) were classified as non-entailers. Among the three non-entailers who emitted an accurate rule, two answered 83% of trials correct on the test of combinatorial entailment, missing the entailer cutoff by only one correct trial. However, the third participant answered only 39% of trials correctly on the test of combinatorial entailment. Of the 56 participants who did not emit accurate rules (i.e., classified as other), 27 (48.21%) of them were non-entailers and 29 were entailers (51.79%).

An eight time point (assessment blocks) by two (accurate or other rule classification) mixed model ANOVA was conducted to assess whether or not report of accurate self-generated rules moderated preference towards coherent contexts. The assumption of sphericity was violated in this model, W = .037, $\chi^2(27)$ = 235.154, $p < .001$, Greenhouse-Geisser ε = .454, and a Greenhouse-Geisser correction was applied to mitigate the violation. In addition, Box's test of equality of covariance matrices was significant, M = 84.922, F(36, 4539.46) = 1.962, p = .001, indicating that the assumption of homoscedasticity was not met. However, the robustness of this model to the assumption of homoscedasticity was assumed as the significance value of the test did not meet the p < .001 alpha criterion recommended by Tabachnick and Fidell (2007).

There was significant interaction of assessment block by rule classification, $F(3.177, 235.122) = 14.904$, $p < .001$, partial $\eta^2 = .168$, and the main effect for assessment block remained significant in the model, $F(3.177, 235.122) = 34.992$, $p < .001$, partial $\eta^2 = .321$. There was not a significant main effect for rule classification, $F(1, 74) = .762$, $p = .385$, partial $\eta^2 = .01$. Figure 17 presents mean percentage of responses allocated towards the coherent contextual cue at each assessment point by rule classification (accurate or other).
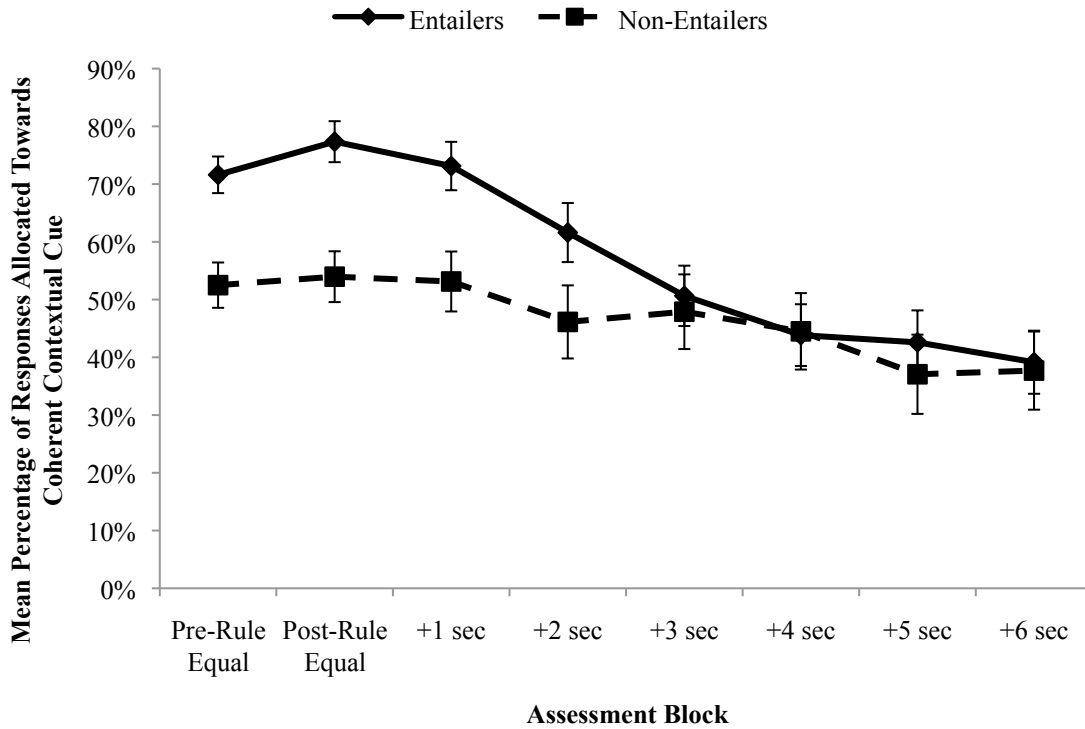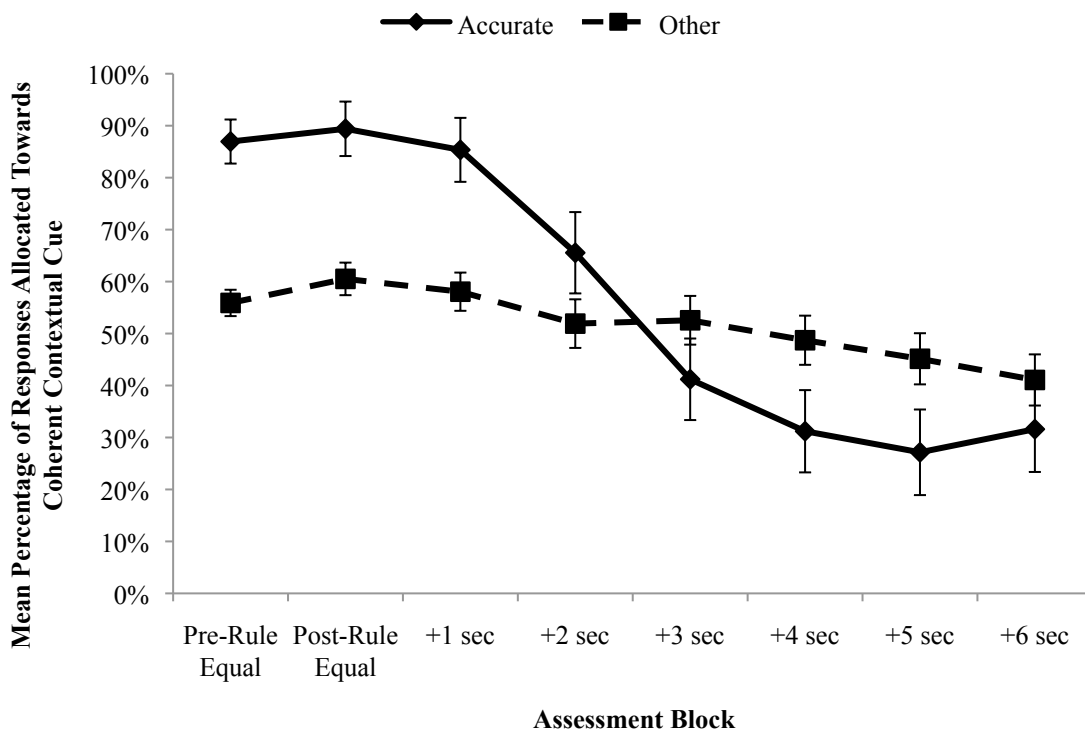

Figure 17. Percentage of responses allocated towards the coherent contextual cue across the eight preference assessment blocks by rule classification.

A follow up analysis was conducted to explore the significant interaction effect of assessment block by rule classification. Independent sample t-tests were used to compare entailer and non-entailer response allocations towards coherence across each of the eight preference assessment blocks. There was a significant difference between participants who generated accurate rules and those who did not during both the pre-rule equal assessment block

t(74) = -6.286, p < .001, *d* =1.68, and post-rule equal assessment block, t(74) = -4.731, p < .001, *d* = 1.42. Participants who generated accurate rules had greater response allocations towards the coherent contextual cue in both the pre-rule equal (M = 86.95, SD = 17.25) and post-rule equal (M = 89.4, SD = 17.25) assessment blocks compared to participants who did not generate accurate rules (pre-rule equal M = 55.89, SD = 19.52; post-rule equal M = 60.52, SD = 23.12).

There was also a significant difference in the +1 second response cost block, t(74) = -3.806, p < .001, *d* = 0.95, with accurate rule generators (M = 85.35, SD = 31.15) displaying greater responding to coherent contextual cue compared to other rule generators (M = 58.05, SD = 26.17). No other comparisons were statistically significant. However, it is noteworthy that participants who generated accurate rules displayed relatively greater responses allocated towards the incoherent cue during later stages of the response cost assessment compared to participants who generated non-accurate rules. While follow up tests of these mean differences did not reach significance, the trend was supported by the presence of a significant cubic contrast for the assessment block by combinatorial entailment performance interaction, F(1, 74) = 20.644, p < .001, partial η^2 = .218. The linear contrast for the interaction term was also significant in the main model, F(1, 74) = 25.353, p < .001, partial η^2 = .255, with ocular inspection revealing a clear decreasing trend across the blocks among participants who generated accurate rules compared to a relatively flat trend among participants who generated other rules.

To assess whether or not participants changed their responding to become more consistent with their self-generated rules, a Bhapkar test of marginal homogeneity was conducted comparing the response pattern classifications of the pre-rule equal assessment block to the response pattern classifications of post-rule equal assessment block. There were no significant changes in classifications across the assessment blocks, Bhapkar χ^2 (2, N = 76) = 1.278, p = .528.

A series of follow up analyses were conducted to see if changes in responses classifications between the first two preference assessment blocks were moderated by combinatorial entailment performance or generation of an accurate rule. There were no significant changes in classifications across the assessment blocks for non-combinatorial entailers, Bhapkar χ^2 (2, N = 30) = 1.034, p = .596 or for entailers, Bhapkar χ^2 (2, N = 46) = 1.398, p = .497. There were also no significant changes in classifications across the assessment blocks for participants with a rule classification as other, Bhapkar χ^2 (2, N = 56) = 2.05, p = .359, or among participants who accurately reported the rule, Bhapkar χ^2 (2, N = 20) = 2.22, p = .329.

Prediction model of coherent preference. A hierarchical regression model was used to assess whether or not self-reported measures of psychological distress, psychological inflexibility, and cognitive fusion predicted preference towards coherent contexts. Preference towards coherence was conceptualized as the degree to which participants persisted in their preferences towards coherence in the face of a response cost. To assess the degree of persistence in preference towards coherence, a hierarchical regression model was conducted using only participants who were classified as displaying a coherent preference during the post rule equal assessment block (n = 34). The predicted variable was set as the numeric value of the first block in which each participant was no longer classified as displaying a coherent preference. This value ranged from one (a participant who immediately switched away from coherent responding on the +1 response cost block) to seven (a participant who persisted in their preference towards coherent responding throughout all six response cost assessment blocks). Social desirability (MCSDS-SF) was entered as the first step of the model and psychological distress (GHQ-12), psychological inflexibility (AAQ-II), and cognitive fusion (CFQ) were entered in as the second step. Both the first step, $F(1, 32) = 2.590$, p =.117, and second step of the model, $F(4, 29) = .997$,

p = .425, were not significant, indicating that the self-report measures did not predict persistence

in preference towards coherence during the response cost assessment phase.

Analysis of Self-Reported Frustration

A 17 time point (F-VAS administrations) repeated measures ANOVA was conducted to

determine if there was an overall time effect for frustration visual analogue scale ratings across

the experimental preparation. Mauchly's test of sphericity was significant within the model, W

$= .000$, $\chi^2 (135) = 1115.639$, $p < .001$, Greenhouse-Geisser $\varepsilon = .233$ indicating a departure from

this assumption. A Greenhouse Geisser correction was applied to mitigate this violation. There

was a significant main effect across the 17 F-VAS administrations, $F(3.736, 280.188) = 66.925$,

$p < .001$, partial $\eta^2 = .472$, indicating significant changes in self-reports of frustration across the

17 administrations. Figure 18 presents mean F-VAS values at each time point along with

standard error.
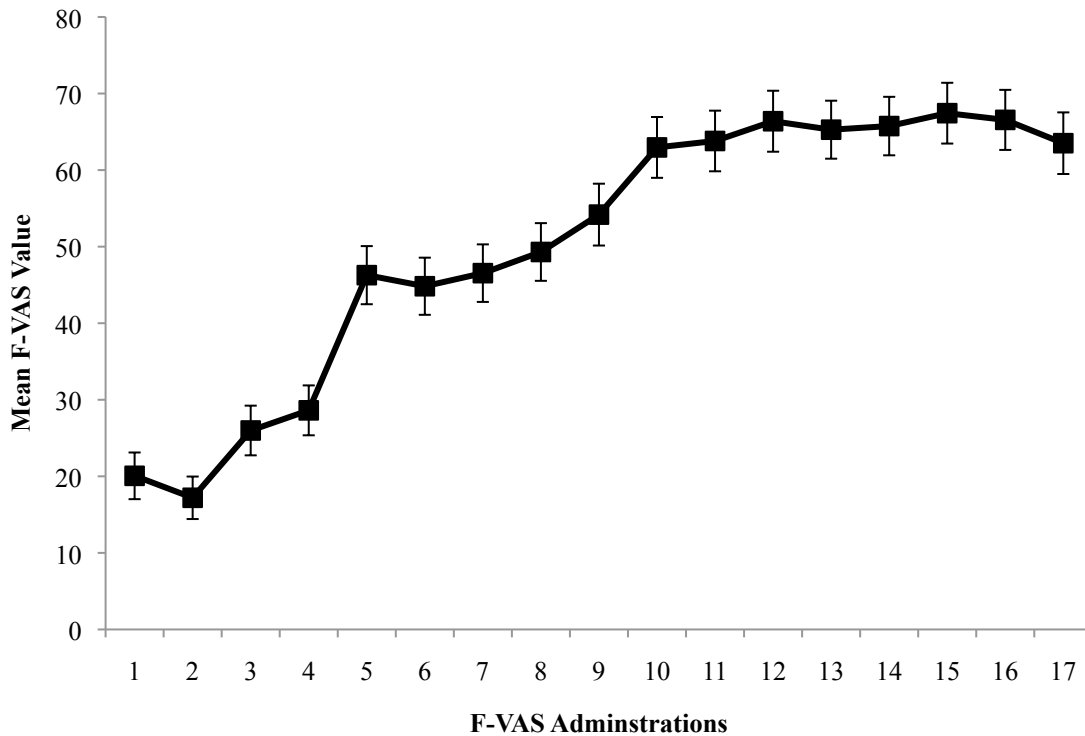
Figure 18. Mean F-VAS scores across the 17 F-VAS administrations.

A follow up polynomial contrast analysis revealed a significant linear, $F(1, 75) = 144.640$, $p < .001$, partial $\eta^2 = .659$, and quadratic, $F(1, 75) = 30.801$, $p < .001$, partial $\eta^2 = .291$, trend in the obtained data. Frustration scores generally increased over time (linear trend) and remained consistently high over later F-VAS assessment time points (quadratic trend). There were also significant 6th order, 7th order, 9th order, 11th order, 12th order, 13th order, 14th order, and 15th order polynomial contrasts. However, these higher order contrasts were not interpreted as they lacked clear conceptual utility to the current analysis. Of particular note in the obtained data is the large increase in self-reported frustration from the 4th administration given at the end of the coherence testing phase ($M = 28.62$, $SD = 28.41$) to 5th administration given after the first part of training in class acquisition phase ($M = 46.28$, $SD = 33.09$), $t(75) = -7.152$, $p < .001$, $d = 0.57$.

Similar, albeit smaller, jumps in frustration were also evident between other study phases. There was a significant difference in frustration between the 2nd administration given at the end of the practice phase (M = 17.20, SD = 24.12) and the 3rd administration given after the first block of the coherence testing phase (M = 25.99, SD = 28.29), t(75) = -4.404, p < .001, d = 0.33. In addition, there was a significant difference between the 9th administration given at the end of the class acquisition phase (M = 54.18, SD = 35.22) and the 10th administration given after the first preference assessment block in the coherence preferences assessment phase (M = 62.96, SD = 34.64), t(75) = -4.091, p < .001, d = 0.25. These observed jumps all coincided with changes between phases of the experimental preparation, indicating that significant increases in frustration were associated with changes in the experimental task.

Moderation Follow Up Analyses. Several planned follow up analyses were conducted to determine if relevant study variables moderated self-reports of frustration within the experimental task. In particular, the potential moderating effects of experimental condition (Meaning or Shapes), performance on the test of combinatorial entailment (Entailers or Non-Entailers), and accurate rule formation during the coherence preference assessment (Accurate or Other) were assessed via three independent mixed model ANOVAs. In addition, the moderating effects of self-report measures (i.e., MCSDS-SF, GHQ, AAQ-II, CFQ,) were assessed via four independent repeated measure ANOVAs containing a continuous covariate predictor.

A 17 time point (F-VAS administration) by two (meaning or shapes condition) mixed model ANOVA was used to determine if experimental condition moderated frustration. Mauchly's test of sphericity was significant, W = .000, $\chi^2(135)$ = 1102.224, p < .001, Greenhouse-Geisser ε = .233 and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. In addition, Box's test of equality of covariance matrices was

significant, $M = 349.712$, F(153, 15382.217) = 1.716, p < .001, indicating that the assumption of homoscedasticity was not met. Given the unequal sample sizes and the fact that the significance value exceeded the p < .001 alpha criterion recommended by Tabachnick and Fidell (2007), the robustness of this model to the assumption of homoscedasticity could not be guaranteed. The main effect for F-VAS administrations remained significant in the model, F(3.739, 276.674) = 64.941, p < .001, partial η^2 = .467. However, there was not a significant main effect for experimental condition, F(1,74) = .001, p = .980, partial η^2 = .000, nor was there a significant condition by F-VAS administrations interaction, F(3.739, 276.674) = .563, p = .678, partial η^2 = .008, indicating that experimental condition did not moderate self reports of frustration. The null hypothesis was retained with confidence, despite the potential violation of the assumption of homoscedasticity, upon the recommendation of Tabachnick and Fidell (2007).

A 17 time point (F-VAS administration) by two (combinatorial entailers or non-combinatorial entailers) mixed model ANOVA was used to determine if performance on the test of combinatorial entailment moderated frustration levels. Mauchly's test of sphericity was significant, $W = .000$, $\chi^2(135) = 1102.444$, p < .001, Greenhouse-Geisser $\varepsilon = .231$ and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. In addition, Box's test of equality of covariance matrices was significant, $M = 330.399$, F(153, 11980.858) = 1.590, p < .001, indicating that the assumption of homoscedasticity was not met. The main effect for F-VAS administrations remained significant in this model, F(3.689, 272.969) = 65.178, p < .001, partial η^2 = .468, and there was not a significant F-VAS administrations by combinatorial entailment performance interaction, F(3.689, 272.969) = 1.178, p = .321, partial η^2 = .016. There was, however, a significant main effect for combinatorial entailment performance, F(1,74) = 4.155, p = .045, partial η^2 = .053, with entailers (M = 45.32, SE 3.87) displaying lower

average levels of frustration compared to non-entailers (M = 57.88, SE =4.80) across the entire preparation (n.b., this observed mean difference should be interpreted with caution as the model's robustness to violations of homoscedasticity cannot be guaranteed).

While the interaction term was not significant in this model, a series of independent t-test follow up analyses were conducted across the 17 F-VAS administrations to explore the significant main effect for combinatorial entailment. Comparisons of mean F-VAS scores by combinatorial entailment performance across the 17 F-VAS administrations is presented in Table 21. The follow up comparisons revealed a significant difference in frustration between entailers and non-entailers only during the F-VAS administrations in the class acquisition phase of the study (i.e., F-VAS administrations 5-9).

Table 21

Mean F-VAS Scores by Combinatorial Entailment Performance Across the 17 F-VAS
Administrations and Comparison by Entailment Status.

F-VAS Administration	Entailers (n = 30)		Non-Entailers (n = 46)		Comparison		
	M	SD	M	SD	t(74)	p	d
1	16.24	25.07	25.93	28.15	1.569	.121	0.36
2	13.59	22.29	22.73	26.28	1.629	.108	0.38
3	24.39	27.36	28.43	29.97	.606	.546	0.14
4	26.30	27.56	32.17	29.78	.878	.383	0.20
5	39.30	28.63	56.97	36.94	2.341	.022*	0.53
6	38.33	29.75	54.80	34.55	2.213	.030*	0.51
7	37.98	28.82	59.67	34.56	2.962	.004*	0.68
8	42.41	30.99	59.87	33.31	2.330	.023*	0.54
9	46.61	33.17	65.80	35.62	2.394	.019*	0.32
10	57.70	33.16	71.33	35.72	1.724	.089	0.40
11	60.50	33.81	68.87	35.56	1.033	.305	0.24
12	62.72	34.79	72.00	34.44	1.142	.257	0.27
13	60.85	32.19	72.07	33.64	1.459	.149	0.34
14	60.65	32.46	73.57	33.58	1.672	.099	0.39
15	62.96	34.87	74.30	33.60	1.406	.164	0.33
16	61.59	33.33	74.20	34.66	1.588	.117	0.37
17	58.46	35.00	71.27	34.34	1.571	.120	0.37

* $p < .05$

A 17 time point (F-VAS administration) by two (accurate or other rule classification) mixed model ANOVA was used to determine whether participants who accurately self-reported the rule during the coherence preference assessment reported different levels of frustration throughout the experimental preparation. Mauchly's test of sphericity was significant, $W = .000$, $\chi^2(135) = 1079.083$, $p < .001$, Greenhouse-Geisser $\varepsilon = .243$ and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. In addition, Box's test of equality of covariance matrices was significant, $M = 446.347$, $F(153, 4208.148) = 1.839$, p

< .001, indicating that the assumption of homoscedasticity was not met. As with the previous analysis, observed mean differences should be interpreted with caution, as this model's robustness to violations of homoscedasticity cannot be guaranteed.

There was a significant main effect for rule classification, $F(1,74) = 6.791$, $p = .011$, partial $\eta^2 = .084$, with accurate rule generators (M = 37.35, SE 5.78) displaying lower average levels of frustration compared to other rule generators (M = 54.89, SE = 3.45) across the entire preparation. There was also a significant rule classification by F-VAS administrations interaction, $F(3.882, 287.239) = 2.556$, $p = .041$, partial $\eta^2 = .033$, indicating that the degree to which accurate rule formation moderated frustration changed over the course of the study.   The main effect for F-VAS administrations remained significant in this model, $F(3.882, 287.239) = 42.973$, $p < .001$, partial $\eta^2 = .367$. Mean F-VAS frustration levels and standard error by rule classification (accurate or other) are presented in Figure 19 to allow for ocular inspection of this significant interaction.
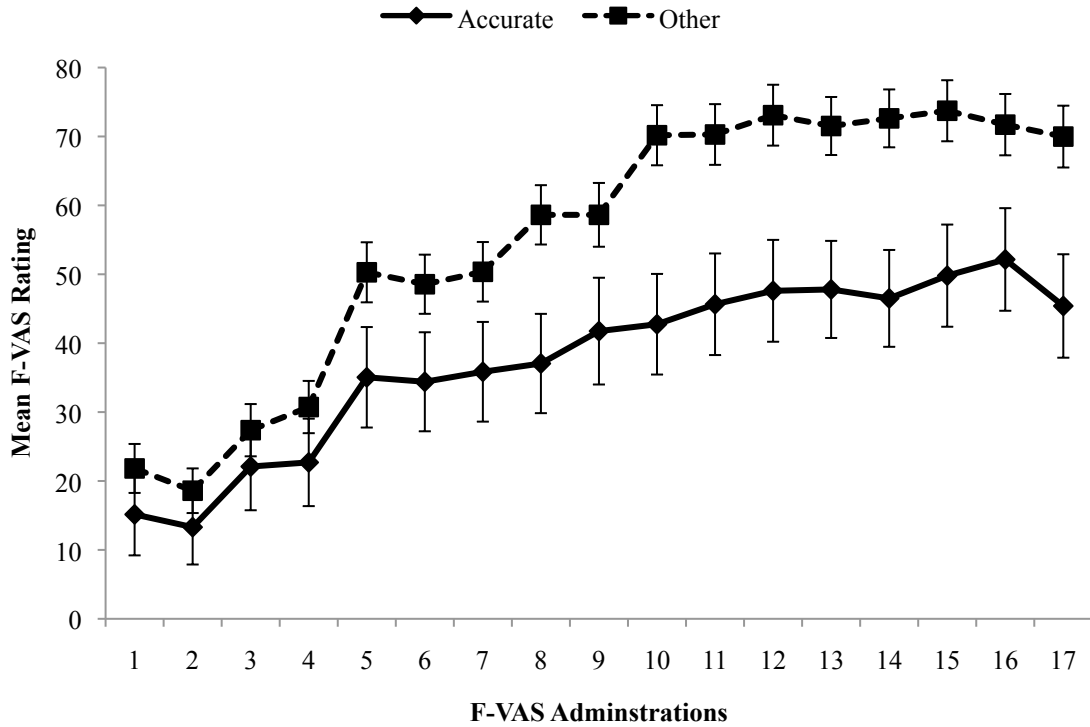
Figure 19. Mean F-VAS scores by rule categorization across the 17 F-VAS administrations.

The linear contrast for the interaction term was significant in the main model, $F(1, 74) =$ 5.061, $p = .027$, partial $\eta^2 = .064$, with ocular inspection revealing a difference in levels of the groups emerging during the 5th F-VAS administrations and growing considerably in magnitude starting in the 10th F-VAS administrations. There was also a significant 9th order polynomial contrast for the rule classification by time interaction, $F(1, 74) = 4.199$, $p = .044$, partial η^2 $= .054$. However, it was not interpreted due to the lack of conceptual clarity. A series of independent t-test follow up analyses were conducted across the 17 F-VAS administrations to determine the source of the significant interaction effect. Comparisons of mean F-VAS scores by rule categorization (accurate or other) across the 17 F-VAS administrations is presented in Table 22. The follow up comparisons indicated a significant departure from equality of frustration ratings between accurate and other rule generators beginning in the 10th F-VAS

administration (given at the end of the pre rule equal coherent preference assessment block) and

persisting for the remainder of the experimental preparation. Accurate rule generators reported

significantly lower levels of frustration than other rule generators only during the eight

assessment blocks of the coherence preference assessment (i.e., F-VAS administrations 10-17).

Table 22

Mean F-VAS Scores by Rule Categorization Across the 17 F-VAS Administrations and
Comparison by Rule Classification.

F-VAS Administration	Accurate Rule (n = 20)		Other Rule (n = 56)		Comparison		
	M	SD	M	SD	t(74)	p	d
1	15.15	24.81	21.82	27.18	0.963	.339	0.26
2	13.30	23.04	18.59	24.62	0.838	.405	0.26
3	22.10	27.22	27.38	28.77	0.714	.478	0.19
4	22.70	26.89	30.73	28.87	1.087	.281	0.29
5	35.05	28.43	50.29	33.94	1.793	.077	0.49
6	34.40	26.36	48.55	33.90	1.691	.095	0.47
7	35.85	31.52	50.36	32.64	1.721	.089	0.45
8	37.05	31.71	53.68	32.40	1.981	.051	0.52
9	41.75	34.29	58.63	34.76	1.870	.065	0.49
10	42.75	35.69	70.18	31.53	3.225	.002*	0.81
11	45.65	33.38	70.29	32.83	2.868	.005*	0.74
12	47.60	32.80	73.09	33.14	2.961	.004*	0.77
13	47.80	32.41	71.52	31.19	2.890	.005*	0.75
14	46.50	32.84	72.63	30.93	3.191	.002*	0.82
15	49.80	38.17	73.73	31.24	2.771	.007*	0.69
16	52.15	37.81	71.71	31.59	2.255	.027*	0.56
17	45.40	34.50	69.98	33.25	2.881	.006*	0.73

* p < .05

The potential moderating effect of social desirability on self-reports of frustration was

evaluated using a 17 time point (F-VAS administrations) repeated measures ANOVA with

MCSDS-SF scores entered as a continuous covariate predictor. Mauchly's test of sphericity was

significant, $W = .000$, $\chi^2(135) = 1085.212$, $p < .001$, Greenhouse-Geisser $\varepsilon = .237$ and a

Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. There

was not a significant interaction between social desirability and F-VAS administrations, $F(3.793, 280.679) = 2.090$, $p = .086$, partial $\eta^2 = .027$, and the main effect for F-VAS administrations remained significant in the model, $F(3.793, 280.679) = 21.712$, $p < .001$, partial $\eta^2 = .227$. There was a significant main effect for social desirability, $F(1, 74) = 11.558$, $p = .001$, partial $\eta^2 = .135$. A follow-up linear regression was conducted to determine the direction of the effect using MCSDS-SF scores as the predictor variable and the sum of each participant's 17 F-VAS scores as the predicted variable. Greater levels of social desirability were associated with lower overall reports of frustration, $\beta = -.368$, $t = -3.40$, $p = .001$.

The relationship between psychological inflexibility and frustration was evaluated using a 17 time point (F-VAS administrations) repeated measures ANOVA with AAQ-II scores entered as a continuous covariate predictor. Mauchly's test of sphericity was significant, $W = .000$, $\chi^2(135) = 1087.567$, $p < .001$, Greenhouse-Geisser $\varepsilon = .237$ and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. There was not a significant interaction between psychological inflexibility and F-VAS administrations, $F(3.790, 280.484) = 1.717$, $p = .150$, partial $\eta^2 = .023$. The main effect for F-VAS administrations remained significant in the model, $F(3.790, 280.484) = 20.820$, $p < .001$, partial $\eta^2 = .220$, and there was a significant main effect for psychological inflexibility, $F(1, 74) = 6.747$, $p = .011$, partial $\eta^2 = .084$. A follow-up linear regression was conducted to determine the direction of the effect using AAQ-II scores as the predictor variable and the sum of each participant's 17 F-VAS scores as the predicted variable. Greater levels of psychological inflexibility were associated with greater overall levels of frustration, $\beta = .298$, $t = 2.597$, $p = .011$. However, a subsequent hierarchical regression model controlling for social desirability during the first step and assessing psychological

flexibility at the second step found that the effect of psychological inflexibility became non-significant after controlling for social desirability, $\beta = .204$, t = 1.840, p = .070.

The relationship between cognitive fusion and frustration was evaluated using a 17 time point (F-VAS administrations) repeated measures ANOVA with CFQ scores entered as a continuous covariate predictor. Mauchly's test of sphericity was significant, $W = .000$, $\chi^2(135) = 1092.637$, p < .001, Greenhouse-Geisser $\varepsilon = .231$ and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. There was not a significant interaction between cognitive fusion and F-VAS administrations, $F(3.694, 273.374) = 2.385$, p = .056, partial $\eta^2 = .031$, and the main effect for F-VAS administrations remained significant in the model, $F(3.694, 273.374) = 8.124$, p < .001, partial $\eta^2 = .099$. There was a significant main effect for cognitive fusion, $F(1, 74) = 10.046$, p = .002, partial $\eta^2 = .120$. A follow-up linear regression was conducted to determine the direction of the effect using CFQ scores as the predictor variable and the sum of each participant's 17 F-VAS scores as the predicted variable. Greater levels of cognitive fusion were associated with greater overall levels of frustration, $\beta = .346$, t = 3.170, p = .002. This effect remained significant even after the effect of social desirability was controlled for in a hierarchical regression model, $\beta = .234$, t = 2.008, p = .048.

Finally, the relationship between general psychological distress and frustration was evaluated using a 17 time point (F-VAS administrations) repeated measures ANOVA with GHQ-12 scores entered as a continuous covariate predictor. Mauchly's test of sphericity was significant, $W = .000$, $\chi^2(135) = 1049.204$, p < .001, Greenhouse-Geisser $\varepsilon = .252$ and a Greenhouse-Geisser correction was applied to mitigate the violation of this assumption. There was a significant interaction between general psychological distress and F-VAS administrations, $F(4.027, 298.033) = 5.042$, p = .001, partial $\eta^2 = .064$, and no overall effect for the GHQ, F(1,

74) = 1.269, p = .264, partial η^2 = .017. The main effect of F-VAS administrations remained significant in the model, F(4.027, 298.033) = 67.175, p < .001, partial η^2 = .476.

The linear contrast of the interaction term was significant in the main model, F(1, 74) = 11.365, p = .001 , partial η^2 = .133. To explore this interaction, a follow up linear regression analysis was conducted using GHQ-12 scores to predict F-VAS levels across each of the 17 F-VAS administrations. The standardized regression coefficient (β) and significance test of the GHQ-12 as a predictor of frustration across all 17 F-VAS administrations are presented in Table 23. The follow up analysis revealed that overall psychological distress as measured by the GHQ-12 significantly predicted higher levels of reported frustration throughout the first five administrations of the F-VAS but did not significantly predict frustration for any of the other F-VAS administrations. That is, baseline levels of psychological distress resulted in greater frustration reports throughout the early experimental tasks but did not significantly predict frustration reports during the middle and later portions of the experimental preparation.

Table 23

General Psychological Distress (GHQ-12) as a Predictor of Frustration (F-VAS) Level Across the 17 F-VAS Administrations

F-VAS Administration	β	t	p
1	.362	3.336	.001*
2	.371	3.440	.001*
3	.417	3.950	<.001*
4	.396	3.707	<.001*
5	.228	2.015	.048*
6	.185	1.619	.110
7	.106	0.914	.364
8	.071	0.616	.540
9	.055	0.471	.639
10	.003	0.026	.979
11	-.048	-0.413	.681
12	.046	0.392	.696
13	.014	0.120	.905
14	-.025	-0.217	.828
15	-.029	-0.254	.801
16	-.057	-.495	.622
17	-.011	-0.097	.923

* p < .05

DISCUSSION

Overall findings from this investigation lend empirical support for the relational frame theory (RFT) assertion that coherence, the act of deriving relational responding in an internally consistent manner, is a well established operant repertoire in verbally capable humans. In addition, the obtained results provide evidence suggestive of the reinforcing properties of coherence among a subset of study participants. Findings and implications from the coherence testing, coherence preference assessment, and frustration analysis will be discussed in turn with consideration given to applied implications at the conclusion of this section.

Coherence Testing

Study hypotheses regarding performance in the coherence testing phase were all supported at least partially by the obtained data. In particular, the hypothesis that slight differences in learning history would result in different patterns of responding on the ambiguous task was fully supported by the obtained findings. The hypothesis that participants would report self-generated verbal rules that are largely consistent with their response patterns was largely supported. In addition, the hypothesis that participants would behave more consistently with their own rules when asked to complete an ambiguous task for a second time was only partially and inconclusively supported.

Departure from random responding. Within both the first and especially within the second testing block, the majority of participants displayed response patterns consistent with either the shape based or meaning based response strategy assessed by the experimenter. This

finding suggests that coherent responding is a well established operant repertoire, as these response patterns emerged without any programmed reinforcement or specific instructions to respond in a coherent or accurate manner.

Furthermore, it is not appropriate to assume that the participants whose response patterns were classified as other responding were engaged in purely random responding during the task. That is, participants who failed to respond according to the meaning based or shape based strategies assessed by the experimenter may have been responding in an internally consistent and coherent manner using a different rule to guide their responding. In fact, 80 of the 81 participants reported using some sort of rule to guide their responding during the task with only one participant reporting that they just chose randomly. For example, several meaning condition participants reported using a hybrid rule that matched round foods with the circle stimulus, square foods with the square stimulus, and all moldy or disgusting foods with the triangle stimulus. Participants who reported following this hybrid rule would be classified as other responding, even though they may have been consistently following the rule throughout the task.

Previous research has found that participants who do not respond in ways anticipated by the experimenter often engage in an internally consistent idiographic response strategy. For example, Holth and Arntzen (1998) found that 15 of 21 subjects who did not respond with the planned equivalence relations display a consistent pattern of responding following another set of relations. In addition, Wilson and Hayes (1996) found that 22 out of 23 participants displayed internally consistent response patterns within a resurgence preparation.

A limitation of the current study is the lack of an internal consistency analysis of all response patterns within each coherence testing block. The experimental preparation did allow for the detection of internally consistent response patterns among participants who followed the

experimenter anticipated response rules (i.e., meaning and shapes based). However, only a small number of participants who displayed other response patterns provided adequate detail in their self-reported rules to allow for post-hoc evaluation of their response strategy. While possible in principle, it was deemed beyond the scope of the current investigation to attempt to infer the presence of coherent and consistent rules from the observed response patterns of participants who engaged in other responding. The nature of the testing task made such an analysis infeasible, as the lack of a programmed requirement to match the three shapes class members equally to the foods stimuli allowed for a large number of possible response rules to be generated. Furthermore, allowing for the possibility of random errors in rule following (e.g., emitting a rule incongruent response on a small number of trials) would have made the inferential analysis more difficult and unacceptably subjective. Future studies should consider using a more elaborate self-report procedure (e.g., a post-task interview or protocol analysis) to ensure each participant's self-reported response strategy contains enough detail enough to allow for an internal consistency analysis of their responding.

Even absent a complete analysis of participants displaying a non-specified response pattern, the finding that the majority of participants displayed response patterns that were internally consistent with anticipated rules has several implications for the design of matching to sample preparations. Many human operant studies employ pre-tests with no programmed reinforcement in order to establish the lack of derived relations or relevant contextual control prior to experimental training and subsequent acquisition. However, most participants in the current study did not respond randomly during testing trials that were presented in a manner consistent with a pre-test. This obtained finding suggests that the underlying assumption of a

108

pre-test, namely that participants will display scores consistent with chance responding on experimenter specified relations, may be violated.

The emergence of rule-based and internally consistent response patterns in the absence of programmed contingencies has been demonstrated before in the literature. For example, Harrison and Green (1990) cautioned that untrained relations can emerge during unreinforced testing trials. Their analysis found that repeated pairings of stimuli in testing arrays can lead to the acquisition of untrained relations. The current findings extend this caution beyond just stimulus arrangements, as responding came under relational stimulus control within a single testing block without any repeated stimulus pairings. Within the first testing block, Foods Class stimuli were presented only once each as a sample, and they were always paired with the three Shapes class stimuli (the arbitrary shapes of lines, circles, and triangles). Thus, the emergence of internally consistent responding in the absence of reinforcement does not appear to be limited to cases of repeated stimuli parings.

Another implication of the obtained findings is that experimenters should be cautious in interpreting matching to sample pre-test results indicative of random responding as actual evidence of random responding. An observation that a participant obtained chance consistent scores on the experimenter specified relations indicates only that the participant did not follow the experimenter specified response pattern. It does not necessarily demonstrate random responding. It could be the case, as it was for many participants in this study who followed either the shape or meaning rule, that the participant was simply following a rule-based and internally consistent response strategy that was different from the one assessed by the experimenter.

Context counts: Antecedent control of sense making. Findings from the present investigation provide evidence of antecedent control of coherence as an operant. The majority of participants in the Meaning condition, who were given exposure to meaning class stimuli prior to the coherence testing blocks, displayed response patterns consistent with meaning-based rules in both blocks of the coherence testing task. In contrast, no participants in the Shapes condition, who were not exposed to the meaning class stimuli, displayed meaning consistent responding during the coherence testing blocks. This finding highlights the fact that a small manipulation in the history of interaction with experimental stimuli can lead to large differences in obtained response patterns.

Some participants in both conditions displayed response patterns consistent with shape based responding. This indicates that coherent and internally consistent response strategies can emerge based on the formal properties of stimuli absent any other history of interaction, programmed reinforcement, or experimental instruction. However, a history of interaction with the meaning based stimuli was necessary for the emergence of coherent and internally consistent response patterns based on the functional properties of the Foods stimuli (i.e., the healthiness of each foods class member). Relational frame theory (RFT) provides an explanation for how this meaning consistent responding might have emerged.

While studies based on RFT often use programmed contingences to facilitate the acquisition of derived relations, a fundamental assumption of the theory is that engagement in derived relational responding occurs naturalistically and spontaneously in the absence of contrived experimental conditions (Hayes, Barnes-Holmes, & Roche, 2001). Within the current study, the emergence of meaning based responding during the coherence testing task is suggestive of a spontaneous display of derived relational responding and transformation of

stimulus functions. Participants who engaged in meaning based responding interacted with the

shapes stimuli as if they were in a frame of equivalence with the meaning based stimuli (e.g.,

"the lines mean healthy, the circles mean unhealthy, and the triangles mean disgusting").

Furthermore, the functional properties of the shapes stimuli appeared to have been transformed

to include the healthiness functions of meaning class (e.g., "healthy foods go with the stimulus

equivalent to healthy", etc.). The display of combinatorial entailed relations of equivalence and

the transformation of functions from the meaning class stimuli to the shapes class stimuli in the

current preparation provide preliminary empirical support for the core RFT assumption that

derived relational responding occurs spontaneously in the absence of directly reinforced class

acquisition.

This finding has further implications for matching to sample preparations that employ a

pre-test design. Caution should be taken in the sequencing of pre-test blocks to ensure that one

set of tested stimulus relations do not inadvertently lead to spontaneous derived relations and

transformation of stimulus function to a subsequent set of tested relations. This phenomenon,

evidenced by the findings above, can lead to the emergence of derived relations during pre-test

prior to any programmed reinforcement or experimental instruction. For example, a pilot study

conducted by this experimenter found that when a pre-test of the functions of arbitrary stimuli

was placed after a pre-test of the functions of familiar stimuli one third of participants displayed

emergence of derived relations consistent with experimenter expectations (Almada, Bordieri,

Wilson, Kellum, & Gregg, 2010). That is, participants responded to the arbitrary stimuli as if

they were in frames of equivalence with the meaningful stimuli. When the order of pre-test was

reversed in a subsequent study, only 1.14% of participants displayed this response pattern

(Bordieri, Flynn, Kellum, & Wilson, 2011). As a whole, these findings suggest that relations

between arbitrary stimuli should be assessed prior to familiar stimuli when sequencing pre-testing blocks.

The impact of rule generation and rule following. Participants were largely accurate when self-reporting verbal rules that guided their responding during the first testing block. This conclusion is drawn primarily from the obtained finding of high consistency between self-reported rules and responding in the first testing block. In addition, the finding that the significant differences in rule classification between experimental conditions largely corresponded to the differences in response classifications between conditions during the first testing block lends further support to this conclusion. These findings are congruent with previous research that has demonstrated that participants can accurately describe their performance on complex operant tasks (Hayes, Thompson, & Hayes, 1989). The obtained findings also suggest that it may have been easier for participants to discriminate the stimuli based on their healthiness functions than on their formal topography (i.e., shape). That is, there were 16 participants who emitted a shapes based rule but did not engage in shaped based responding compared to only one participant who emitted a meaning based rule but did not engage in meaning based responding. It could be the case that the shapes rule was harder to follow as many of the foods class stimuli shared only vague topographical similarity with their respective shapes class stimuli. In contrast, the healthiness functions of the stimuli may have been more readily apparent (e.g., the presence or absence of mold, pizza compared to fruit) and easier to discriminate. It could also be the case that the meaning functions of food stimuli competed with attention to the topographical properties that might have otherwise organized response classes.

While overall consistency between response patterns and self-reported rules was high, participants in the Shapes condition displayed significantly lower consistency compared to participants in the Meaning condition. It could be the case that this observed difference in consistency is related to the relatively greater difficulty in following the shapes based rule. However, this observed difference might be the result of a methodological artifact of the study design. Participants in the meaning condition gained exposure to all 27 foods class members prior to engaging in the coherence testing task while Shapes condition participants immediately began the testing task without having the benefit of any exposure to the foods class members. Consequently, meaning condition participants acquired familiarity with the foods stimuli prior to the testing task, which may have allowed them to begin following a rule based strategy on the first trial of the task. In contrast, shapes condition participants may have had a more difficult time forming rule based responses during early trials in the testing block.

This limitation in the study design also confounds the analysis of changes in response patterns between the first testing block (before the self-report of the rule) and the second testing block (after the self-report of the rule). While global changes in response patterns were noted between the two blocks, it is unclear whether the emitting of self-generated rules was responsible for the increase in meaning and shape consistent responding observed in the second testing block. Given that the majority of response pattern changes occurred among shapes condition participants, it is likely the case that both the reporting of self-generated rules and the additional opportunity to engage in the testing task with increased foods class stimuli familiarity were contributors to the observed increase in response patterns consistent with self-reported rules. While the mechanism of the change between the testing blocks remains inconclusive, the direction of the change is not. All but two of the 15 participants who changed in response

classifications between the two testing blocks moved from other responding to either a meaning or shapes consistent response pattern. Thus, the self-reporting of rules and the opportunity to reengage in the testing task were likely responsible for an increase in coherent and internally consistent responding, though the specific mechanism responsible remains unidentified.

Future studies should control for familiarity by ensuring both experimental conditions have equal exposure to the stimuli prior to engaging in the testing task. For example, shapes condition participants could be exposed to an identity matching task where they are asked to select the sample foods class stimulus from an array containing the sample stimulus and two other randomly selected foods class members. This task would allow for participants to gain familiarity with stimuli without introducing the possibility of other confounding sources of stimulus control. In addition, future studies should seek to provide a more time sensitive measure of self-generated rules. That is, even if the familiarity confound was controlled for in the current study, it would not be possible to parse the effects of self-reported rules from the mere practice effect inherent in repeating the testing task for a second time. One possible solution to the confounding of practice and rule reporting would be to conduct a protocol analysis in which participants are asked to talk aloud and explain their responding during the task. Another possibility would be to instruct participants to press a button when they have figured out the task and then subsequently obtain a self-report of their generated rules. Using either procedure, responses emitted prior to the formation of the rule could be compared to responses emitted after to determine whether or not changes in consistency of responding occurred concomitantly with the formation of a rule.

Possible maintaining consequences of the observed responding. Overall findings from the coherence testing phase indicate the presence of internally consistent and coherent

responding generated in the absence of programmed contingences or experimental instruction. This begs the question of why such responding occurred. Given the nature of the task, a random responding response strategy (i.e., repeatedly clicking the same stimulus position in the matching array without regards to any of the stimuli features) would have required considerably less effort to engage in than the coherent response patterns obtained. As such, the substantial presence of coherent and internally consistent responding with this testing task is suggestive of the reinforcing nature of coherence.

One possible limitation to the above conclusion was the fact that participants were told during the informed consent procedures that the study is, "interested in investigating the ways in which people make sense of ambiguous tasks." Thus, while care was taken to remove all standard instructions regarding accuracy and the presence of correct answers during the matching to sample task instructions, it is possible that the informed consent procedure itself may have served as a distal experimental instruction that influenced the observed findings. Future studies should consider employing institutional review board approved deception procedures to ensure that participants are not exposed to any references to sense making prior to the debriefing given after study completion.

Another possible limitation was the presence of differences in social desirability between experimental conditions. Participants with greater levels of socially desirable response tendencies are assumed to possess a greater desire to please the experimenter (Nederhof, 1985). It is possible that they may have responded in a more coherent manner on the testing task as a result. However, participants in the shapes condition displayed both higher levels of social desirability and lower levels of responding consistent with the experimenter anticipated response

strategies. This suggests that group differences in social desirability were not a major threat to the obtained findings.

Coherence Preference Assessment

Some study hypotheses regarding performance in the coherence preference assessment and coherence preference assessment with response cost phases of the study were supported by the obtained data while others were not. The hypothesis that participants would display a preference towards the coherent context during the equal preference assessment task was fully supported. Support was also obtained for the prediction that participants would, to varying degrees, persist in preference towards coherent contexts when an aversive consequence is in place. However, all hypothesized relationships between persistence in preference towards the coherent context and measures of psychological inflexibility, cognitive fusion, and psychological distress were not supported.

Global preference towards coherence and subsequent changes in preference. Taken as a whole, participant responding during the two equal preference assessment blocks was indicative of a preference towards contexts where coherent responding was possible compared to contexts that could not be solved in a way consistent with the class acquisition training. On the group level, the mean percentage of responding allocated towards the coherent contextual cue was significantly greater than chance would predict for both equal preference assessment blocks. However, claims of robust evidence of coherence preference are tempered by the single subject response pattern analysis, which revealed a weaker effect (i.e., only 39.47% and 44.74% of participants displayed a clear preference towards the coherent context during the pre rule equal and post rule equal blocks, respectively).

These findings are largely consistent with those obtained by Wray (2011; see also Wray, Dougher, Hamilton, & Guinther, 2012) despite the use of a different methodology. Wray and colleagues provided instructions to participants that encouraged problem solving and sense making during the solvable conditions. However, the preference assessment task in the current study was conducted without any direct instructions given to participants regarding how to respond. In addition, while Wray and colleagues used a yoked reinforcement procedure to provide programmed reinforcement in both the solvable and unsolvable conditions, no programmed reinforcement was provided during the preference assessment in the current study. Thus, the obtained findings both replicate and extend Wray and colleagues' work by displaying the emergence of preference towards coherent contexts among many study participants under different experimental conditions.

There was a global trend away from preference towards the coherent context beginning after the +1 second response cost block and culminating with a significant preference away from the coherent contextual cue during the +5 second and +6 second response cost blocks. This finding supports the study hypothesis that most participants would switch away from a preference towards coherent contexts when it became increasingly costly. However, a small subset of participants (n = 12, 15.79%) persisted in their preference towards the coherent context across all six response cost blocks. These individuals are of particular theoretical interest as their responding is suggestive of maladaptive rule following in the face of direct aversive contingencies. Further consideration of these responders will be given during the concluding discussion of applied implications.

Social desirability moderated preference towards coherent contexts throughout the eight blocks of the coherence preference assessment with greater endorsement of socially desirable

responses associated with a greater global preference towards coherent contexts. Traditionally, social desirability is conceptualized as a threat to the validity that must be contained or otherwise mitigated to prevent contamination and confounding of findings (Nederhof, 1985). Following this logic, the presence of a social desirability effect would indicate a potential validity problem with the preference assessment task.

However, it is worth considering the reasons why greater responding towards the coherent context would be associated with social desirability. A core assumption of the RFT account of coherence holds that coherent relational responding is initially acquired via a rich history of social mediated reinforcement (Hayes, Strosahl, & Wilson, 2011, p. 51). Thus, instead of being considered a threat to the assessment of coherence, the presence of a significant relationship between social desirability and preference towards coherent contexts is supportive of the very contingencies theoretically speculated to be responsible for the acquisition of the behavior of interest. Responding in a coherent fashion is theorized to be acquired due to socially mediated reinforcement. Thus, individuals who have a greater sensitivity to socially mediated reinforcement would be expected to be more likely to engage in a wide array of behaviors associated with increased likelihood of social praise, including engagement in coherent relational responding.

To further explore the nature of this relationship, future studies should be conducted to directly test whether or not socially mediated reinforcement influences preference towards the coherent context. For example, a study could be conducted whereby programmed reinforcement (e.g., "Correct") is provided for choosing any response during incoherent trials while responding correctly on coherent trials results in either no programmed consequence or a programmed punisher (e.g., "Wrong"). Such a preparation would be able to parse whether participants are

more sensitive to socially mediated reinforcement provided within the incoherent contextual cue or to other sources of reinforcement that may be available by allocating responding towards the coherent contextual cue.

Role of combinatorial entailment. The failure to find preference towards the coherence context among the majority of study participants is in need of further exploration. It could have been the case that forced exposure to the incoherent contexts during the forced choice blocks preceding each equal preference assessment block interfered with the derived relations established during class acquisition. Devany, Hayes, and Nelson (1986) speculated that the interspersing of unsolvable trials throughout a matching to sample procedure may interfere with the acquisition and maintenance of equivalence relations. Thus, forced exposure to unsolvable trials immediately prior to the preference assessment may have interfered with the derived relations established during class acquisition for some participants. However, given established findings that equivalence classes are relatively robust to disruption in verbally competent adults, it is unlikely that the limited forced exposure to incoherence contexts was a major source of disruption in this study (Saunders, Saunders, Kirby & Spradlin, 1988; Spradlin, Saunders, & Saunders, 1992).

A more likely explanation is that only participants who acquired both mutual and combinatorial entailed relations of equivalence were able to fully discriminate the differences between the coherent and incoherent contexts. Each matching to sample testing trial in the coherent terminal link had an equal chance of being either a probe of mutual entailment or a probe of combinatorial entailment. Almost all participants who displayed combinatorial entailment also displayed mutual entailment (n = 42, 91.3%). This suggests that combinatorial entailment performance alone is a valid indicator of whether participants had the relational

119

repertoire necessary to fully discriminate between coherent and incoherent contexts. If a participant failed to acquire combinatorial entailment, the functional distinction between coherent and incoherent contexts would at most be only partially detectable. That is, this acquisition failure would result in the functional property of coherence being present at most during only half of the matching to sample trials in the coherent terminal link (i.e., only during mutual entailment probes). Of the 30 participants who failed to display combinatorial entailment, 19 displayed mutual entailment. These non-combinatorial entailers may have been able to discriminate that the coherent contextual cue sometimes led to coherent contexts. In contrast, the 11 non-combinatorial entailers who also failed to display mutual entailment were likely to have experienced the coherent and incoherent contexts as equally incoherent.

It is theoretically assumed that non-combinatorial entailers would display a relatively weaker preference towards the coherent context compared to combinatorial entailers, who were in a position to fully discriminate the functional distinction between the incoherent and coherent contexts. Analyses of performance moderated by combinatorial entailment are supportive of this interpretation. Among participants who displayed evidence of combinatorial entailment, the mean number of responses allocated towards the coherent contextual cue was significantly higher than non-entailers during both equal preference assessment blocks and the +1 second response cost bock. This interpretation is further supported by the response pattern analyses, which revealed a significantly greater number of participants displaying a preference towards the coherent contextual cue among entailers compared to non-entailers across the same three assessment blocks. Finally, the finding that non-entailers displayed response patterns during both equal preference assessment blocks that were closer to random responding than entailers

provides further support for the interpretation that non-entailers were not able to fully discriminate the functional distinctions between the coherent and incoherent contexts.

The assumption that a display of combinatorial entailment is necessary to fully discriminate between the contexts was largely supported by the obtained findings. However, it is tempered by the observed presence of a small number of non-combinatorial entailers who displayed a preference towards the coherent contextual cue. It is most likely the case that this responding was due to the presence of a partial discrimination between the contextual cues as discussed above (i.e., non-combinatorial entailers may have still been able to discriminate that the coherent cue sometimes led to coherent contexts). There are several other possible explanations for this observed responding. The absence of a passing score on the test of combinatorial entailment does not necessarily indicate that entailment was absent. That is, a participant may have acquired the combinatorial entailed relations but made a few random errors on the test, resulting in a classification as a non-entailer. This was particularly likely to have occurred in the current study as the test of combinatorial entailment contained only 18 trials, with three or more errors resulting in classification as a non-entailer.

In addition, it may have been the case that non-entailers were attending to the familiarity of the stimuli arrangements presented in the coherent and incoherent trials. Coherent trials consisted of stimuli arrangements that were already presented during the test of combinatorial entailment while incoherent trials consisted of novel stimuli arrangements. Thus, it is also possible that some non-entailers allocated responding based on the familiarity of the coherent stimuli arrangements, resulting in preference towards the coherent context. A future study could control for this confound by comparing the preference of participants who are exposed to class

121

acquisition training to participants who have no training or history with the testing stimuli prior to the preference assessment.

The design of the experimental preparation may have also contributed to some non-entailers displaying a preference towards the coherent contextual cue. Within the design, the topographical features of the cues (i.e., yellow or blue) were randomized across participants to counterbalance the relationship between the color and the functional property of the cue (i.e., whether it lead to coherent or incoherent contexts). While this counterbalancing mitigated the confounding of the features on the group level, it did not do so on the level of the individual participant. That is, within each subject, the topographic properties (i.e., yellow or blue) and functional properties (i.e., access to coherent or incoherent contexts) of the cues were fixed. This arrangement made it impossible to parse preference based on the topographic versus functional properties on the participant level. Consequently, individual classifications of preference towards the coherent cue may have been the result of a preference towards the color or the functional properties of the respective cue. This limitation tempers the conclusions that can be drawn from the response pattern analyses of participants and also provides a plausible explanation as to why some non-entailers displayed a preference towards the coherent contextual cue. Future studies should seek to control for the confounding of the topographic and functional properties of the contextual cues at the individual level. For example, a design where the link between the contextual cues and the coherent/incoherent contexts is switched halfway through the preference assessment would allow for an analysis of whether or not participants were responding to the functional or topographical features of the contextual cues.

Despite this limitation, analysis and interpretation of performance during the preference assessment is still warranted. Over the course of the six response cost blocks, the initial

differences between entailers and non-entailers largely remitted. While not fully evident from the group level analysis, the single subject response pattern analysis revealed that non-entailers were slower to engage in responding towards the incoherent contextual cue compared to entailers. That is, entailers displayed a greater proportion of both coherent and incoherent cue responding during the middle blocks of the response cost assessment (i.e., within the +3 and +4 second response cost blocks). This finding is suggestive of a different stream of contingencies organizing the responding of entailers and non-entailers. The majority of entailers appeared to follow a rule based response strategy, allocating responding towards the coherent cue until it became too costly, at which point they quickly switched over to responding towards the incoherent and less costly cue. In contrast, the responses of non-entailers may have been controlled by contingency shaped behavior, with primarily undifferentiated responding during early blocks and then a gradual trend towards allocating responses to the less costly incoherent cue over the later assessment blocks.

However, this interpretation is inconsistent with established research that has demonstrated that rule-governed behavior is less sensitive to changes in contingences relative to contingency governed behavior (Galizio, 1979; Hayes, Brownstein, Zettle, Rosenfard, & Korn, 1986; Shimoff, Matthew, & Catania, 1986). That is, prior research supports a prediction that non-entailers would be more sensitive to increases in the response cost associated with the coherent contextual cue and consequently allocate their responding more quickly towards the incoherent cue as the response cost increased. However, the obtained findings indicate the opposite, with more non-entailers than entailers engaged in undifferentiated responding during the +3 and +4 second response cost assessment blocks. One possible explanation for this finding is that non-entailers were not primarily engaged in contingency shaped behavior, and instead

engaged in rule-governed behavior of a different sort than entailers. An analysis of self-reported rules of non-entailers indicated that many endorsed a strategy of just randomly guessing (e.g., "I just chose whatever color was closest to my mouse") that may have lead them to not attend to the underlying differences in the contingency. That is, non-entailers may have followed a random responding rule that resulted in greater insensitivity to the increasing response cost associated with the coherent contextual cue.

Factors that may have contributed to low rates of combinatorial entailment. In addition to the exploration of response difference among entailers and non-entailers discussed above, it is worth considering why only 60.5% of participants displayed robust evidence of equivalence relations during the test of combinatorial entailment. The pass rates for tests of combinatorial entailment vary considerably within the established literature, suggesting that contextual features of the class acquisition training can greatly impact the acquisition of combinatorial entailed relations. For example, Drake and Wilson (2008) demonstrated that the presence of instructions clearly linking accurate responding to less time in the task and full compensation (i.e., research credit) resulted in significant increases in the completion rates and test performance of participants. The absence of any such instructions in the current preparation may have been a factor in the low level of combinatorial entailment observed.

In addition, several researchers have found that the familiarly of the stimuli used in the matching to sample task can dramatically influence acquisition rates of combinatorial entailed relations. For example, the use of all Greek letters in a three by three class formation yielded below chance acquisition of equivalence relations (Holth & Arntzen, 1998). Fields, Arntzen, Nartey, and Eillifsen (2012) found that the inclusion of a meaningful stimulus class (i.e., pictures of familiar objects) greatly increased the acquisition of equivalence relations compared to classes

consisting only of ambiguous shapes. Given that all three stimulus classes used in the current preparation consisted of arbitrary stimuli, it may be the case that the lack of a familiar stimulus class resulted in increased difficulty in the formation of combinatorial entailed equivalence relations.

It could also be the case that fatigue was a factor in the observed level of combinatorial entailment performance. The class acquisition task occurred towards the middle of an extended experimental preparation, and participants may have been less attentive and fatigued by the time they started class acquisition training and testing. Regardless of the source of low combinatorial entailment performance, it is of note that performance on the combinatorial entailment test was predicted by the number of trial blocks need to reach criterion during the first phase of class acquisition training (i.e., the A-B training phase). This finding suggests that it may be possible to identify and intervene on participants who are at increased risk to fail the combinatorial entailment during early stages of the training process. For example, future studies should consider providing an instructional prompt or other form of intervention to participants who do not meet criterion after the third trial block of A-B training (i.e., the median trial block participants required to meet criterion in the current study).

Accurate rule generation and making sense. A relatively low proportion of subjects (n = 20, 26.32%) emitted self-generated rules that accurately discriminated the functional properties of the contextual cues (i.e., that one lead to coherent contexts and the other lead to incoherent contexts). Explanations for this finding relate to the reasons for the low rates of combinatorial entailment previously discussed and will not be readdressed here. The display of robust evidence of combinatorial entailment was not sufficient to accurately emit the rule but it does appear to be necessary, except in a very small number of cases. In particular two participants

125

who missed the cut off for entailment status by one error accurately emitted the rule, suggesting that they acquired the entailed equivalence relations but made an additional error on the combinatorial entailment test. However, one participant displayed chance levels of responding on combinatorial entailment but accurately emitted the rule. It could be the case that this participant simply engaged in random responding during the test of entailment even though they had acquired relations; however, a precise explanation for this finding is not apparent. With the exception of this limiting case, these findings provide further evidence to support the theoretical assumption that the combinatorial entailment of equivalence relations was necessary to fully discriminate the functional difference between the contextual cues.

Accurate rule generators displayed different response allocations across the eight preference assessment blocks compared to participants who did not generate an accurate rule. The obtained findings present a clearer and more differentiated picture than the response allocations of entailer and non-entailers previously discussed. On average, accurate rule generators displayed evidence of rule-governed responding with greater allocation of responding towards the coherent contextual cue during early assessment blocks and a rapid change in preference away from coherence during later assessment blocks when the response cost for coherence was high. This response allocation trend stands in contrast to the non-accurate rule generators who, on average, displayed an undifferentiated allocation during early assessment blocks followed by a small and gradual decrease in responding towards the coherent cue. These obtained findings are suggestive of the presence of a rule-based response strategy employed by accurate rule generators that was different from the contingences that guided non-accurate rule generators.

Significant changes in response allocations were not observed after the self-report of rules during the coherence preference assessment. This finding stands in contrast with the significant changes observed after the self-report of rules during the coherence testing phase of the study. One possible explanation for this discrepancy is the relatively lower proportion of participants who generated accurate rules in the preference assessment (26.32%) compared to participants who generated shape or meaning based rules in coherence testing (75.31%). More participants in the coherence testing phase were in a position to alter their responding in ways that would result in changes in classification on the post-rule task. However, the confound of familiarity with stimuli during coherence testing prevents meaningful conclusions regarding differential effects of rule following from being drawn.

Failure to predict persistence in preference. In contrast to the current study's hypotheses, self-reported measures of psychological inflexibility, cognitive fusion, social desirability, and general psychological distress did not predict persistence in preference towards the coherent context. There are several explanations as to why this incongruence between the theoretical account and obtained data occurred.

Only a subset of the sample (n = 34, 44.74%) met criterion for inclusion in the persistence analysis. Thus, it could be the case that there was insufficient power to detect a significant effect. Another possible explanation for the obtained null findings was the relatively low proportion of distressed participants in the sample. Only 18 study participants (23.68%) were above the distress cutoff score on a measure of general psychological distress (GHQ-12). Given the clinical nature of this theoretical prediction, it could be the case that the relationship between psychological inflexibility, cognitive fusion, and general psychological distress may only emerge within a clinical sample that presents with a greater overall level of distress. Future

127

studies should employ a large clinical sample to test these theoretical predictions. A replication of this preparation within a clinical sample would also allow for the comparison of response patterns displayed by distressed participants to the current patterns obtained from a convenience sample of college students.

Other possible explanations for the failure to predict persistence are related to the design of the preparation. The current study used only arbitrary stimuli during the preference assessment and it could be the case that the predicted effect would have occurred if personally meaningful stimuli with high emotional valance were used (e.g., stimuli related to self identity or worry related content). Future studies should consider using such stimuli to provide a more clinically relevant analogue of problematic sense making. The current study design also used forced choice exposure trials to each contextual cue throughout all eight phases of the preference assessment. That is, participants were directly exposed to increasing difference in response costs between the two cues at the beginning of each assessment block. Given that one of the defining features of rule governed behavior is a relative insensitivity to direct contingencies (Hayes, 1989), it could be the case that the current study provided an inappropriate analogue of problematic rule governance. Different response patterns might have emerged if participants were not forced into direct contact with the relative changes in response cost. Future studies should consider removing the forced choice trials during the response cost assessment to provide a more appropriate experimental analogue of rule governance (i.e., one in which the only way for a participant to discover the relative differences in response costs is via the selection of the incoherent cue during a concurrent choice trial).

Finally, it could be the case that there is no relationship between persistence in preference towards coherent contexts and measures of psychological inflexibility, cognitive fusion, and

general psychological distress. While theoretically asserted both in this study and by Wray and colleagues (2012), there may in fact be no relationship. If the null findings obtained in the current study are subsequently and repeatedly replicated, a major theoretical assertion within acceptance and commitment therapy (ACT) would be called into question. However, it is premature to draw such a conclusion given the obtained null findings from only this study.

Is Coherence a reinforcer? Overall findings from the current study provide evidence that access to coherent contexts was preferred over access to incoherent contexts for many participants. Furthermore, stronger preference was found among participants who displayed combinatorial entailment and among participants who generated a rule that accurately discriminated that the contextual cues lead to coherent and incoherent contexts. However, evidence of preference is merely suggestive of and does not provide direct evidence of the reinforcing properties of making sense. Despite Wary and colleagues' (2012) assertion that they found preliminary evidence of sense making functioning as a reinforcer, their preparation also employed only measures of preference. To date, no published studies have established direct evidence of the reinforcing properties of sense making.

The failure to directly assess the reinforcing properties of sense making is a limitation of the current study. Future studies should be designed to include both measures of preference and a direct reinforcer assessment of sense making. For example, moving the differential response costs from the terminal link inter-trial-interval (ITI) to the initial link ITI would be one method to establish a more direct assessment of the reinforcing functions of coherence. That is, instead of testing whether participants were willing to wait longer periods of time between concurrent choice trials as done in the current study, a future study could test whether participants were

willing to wait longer periods of time within a concurrent choice trial to gain access to coherent contexts.

A more direct method of assessing of the reinforcing properties of coherence could be obtained by changing the response schedules on the initial link stimuli. Instead of keeping both consistent at fixed ratio one (FR1) and employing different ITI lengths, a preparation could hold ITI lengths constant and then parametrically increase the ratio scheduled on the coherent initial link while holding the incoherent initial link ratio schedule constant at FR1. Such a preparation would allow for a direct assessment of whether or not participants are willing to work (i.e., fulfill an increasing fixed ratio requirement) to gain access to coherent contexts relative to concurrently available and low response effort access to incoherent contexts.

Frustration and Coherence

Experimental hypotheses regarding the moderators of frustration during the experimental task were fully supported by the obtained findings. The prediction that psychological inflexibility and cognitive fusion would moderate self-reported levels of frustration was fully supported. In addition, several orderly patterns in frustration that emerged during analysis are worthy of discussion.

Changes in frustration during ambiguous tasks. The observed global increase in reports of frustration during the experimental preparation provides evidence that the tasks in the experiment were frustrating for most participants. Multiple causes may be responsible for this finding. For example, vigilance tasks requiring sustained attention for extended periods of time have been demonstrated to result in increased distress and frustration (Warm, Parasuraman, & Matthews, 2008). Likewise, an extensive psychological literature exists supporting the

130

relationship between ambiguity and distress (Cohen, Stotland, & Wolfe, 1955; Sosnowski, 1983; Sosnowski, 1988).

The linear trend of increased frustration observed in this study suggests that increased time in the task led to increased frustration. However, large jumps in level of frustration between phases of the experiment are also supportive of the relationship between ambiguity and frustration. That is, the statistically significant increases in frustration found between the practice testing phase and coherence testing phase, the coherence testing phase and the class acquisition phase, and the class acquisition phase and coherence preference assessment phase, are all suggestive of the role of ambiguity in frustration. When participants were exposed to a new task without clear experimental instruction, their level of frustration consistently increased.

One possible limitation to these obtained findings was that task length was confounded with ambiguity. This was especially problematic during the coherence preference assessment as participants were directly told how many trials were remaining during each concurrent choice and forced choice trial. Thus, the current study cannot conclusively demonstrate that ambiguity was primarily responsible for increased frustration. Future studies should consider interspersing non-ambiguous tasks throughout the experimental preparation to determine whether or not frustration returns to baseline levels after exposure to a simple and non-ambiguous task. For example, the practice test in the current study (i.e., matching colors to names of colors) resulted in reduced levels of self-reported frustration and may be an appropriately clear and non-frustrating task to repeatedly administer in future studies. By interspersing the color matching practice test with ambiguous tasks throughout the preparation, experimental control of frustration could be evaluated using single subject design logic (i.e., an ABAB return to baseline design).

The effect of task performance on reported frustration. Further evidence suggestive of the relationship between ambiguity and frustration is found within moderation analyses of task performance. Participants who displayed combinatorial entailment reported lower average levels of frustration throughout the experiment compared to non-entailers. Likewise, participants who emitted an accurate rule during the coherence preference assessment also reported lower average levels of frustration compared to those who did not emit an accurate rule. These findings suggest that making sense of the task (i.e., responding in accordance with the experimenter programmed contingences and accurately discriminating the functional properties of the contextual cues) results in lower frustration. Put simply, participants who "got it" were less frustrated.

While these effects were large enough to be statistically significant in the global analytic model (i.e., the total sum of all reports of frustration), follow up analyses revealed that the salutary effects of making sense only appeared within the task where sense was made. That is, entailers reported significantly less frustration relative to non-entailers only during the class acquisition phase of the study. Likewise, accurate rule generators reported significantly less frustration relative to other rule generators only during the coherence preference assessment phases of the study. These findings support the theoretical assertion that ambiguity is aversive, as participants who made sense of a particular task only experienced reduced frustration during that particular task and not during other study phases that preceded and/or followed it.

Self-report measures that moderate frustration. Self-report measures administered prior to the ambiguous behavioral tasks also moderated reports of frustration throughout the study. In particular, social desirability, psychological flexibility, cognitive fusion, and general psychology distress all appear to influence frustration during ambiguous tasks. The following discussion will consider each measure in turn.

The observed association between greater levels of socially desirable responses and lower reports of frustration is not surprising, as individuals who wish to present themselves in a positive light would likely not endorse experiencing high levels of frustration during an experimental task (Crowne & Marlowe, 1960; Nederhof, 1985). This observed effect is a limitation of the current analysis of frustration and future studies should seek to employ measures of frustration that are less susceptible to social desirability. For example, galvanic skin response and blood pressure volume have been shown to be sensitive to increases in frustration during a computerized task (Fernandez & Picard, 1998). Future investigations of frustration during human operant tasks should consider employing both self-report and physiological based measures of frustration to provide a more complete accounting of frustration.

Higher levels of cognitive fusion and psychological inflexibility were both associated with higher levels of frustration throughout the study. These findings are congruent with the psychological flexibility model that underlies acceptance and commitment therapy (ACT; Hayes, Stroshal, & Wilson, 2011). In particular, the obtained findings lend support to the theoretical assertion that individuals high in fusion and inflexibility are more sensitive to the aversiveness of ambiguity. Participants who scored high on these self report measures endorsed with greater frequency items such as "I over-analyze situations to the point where it's unhelpful to me" (CFQ item 4) and "worries get in the way of my success" (AAQ-II item 7). Over-analyzing (rumination) and worrying are core features of psychopathology (Bhur & Dugas, 202; Lyubomirsky & Nolen-Hoeskema, 1995; Papageorgiou & Wells, 2003), and are also both suggestive of attempts to resolve ambiguous situations.

The observed effect for psychological inflexibility did not remain statistically significant after controlling for social desirability; however, the effect for cognitive fusion did remain

133

significant. This finding suggests that cognitive fusion (as measured by the CFQ) may be a more powerful predictor of frustration in the face of ambiguity compared to psychological inflexibility (as measured by the AAQ-II). However, this conclusion is tempered by the questionable internal consistency of the CFQ in the current study. These limitations suggest that further refinement of self-report measures may be needed to assess the moderating effects of fusion and inflexibility in ambiguous behavioral tasks.

Participants who reported high levels of psychological distress displayed higher levels of frustration during early tasks in the study but became indistinguishable from less distressed individuals during middle and latter phases of the study. One possible explanation for this observed finding is that individuals high in psychological distress were more prone to frustration throughout the experimental study but that a significant global effect was not found due to a ceiling effect in the frustration measure. That is, high distressed individuals reported high levels of frustration initially and then had no means of differentiating their increased level of frustration during later stages of the study as they already were responding near the ceiling of the F-VAS scale. While it could be the case that highly distressed individuals were simply more prone to frustration during early experimental tasks and not during later tasks, this explanation contradicts well-established findings in the burnout and workplace stress literature. In particular, findings that show a positive relationship between high initial distress and worse performance outcomes, such as job-related frustration, over time (Maslach, Schaufeli, & Leiter, 2001). Future studies should consider alternative measures of frustration, such as the physiological assessment techniques discussed above, to minimize potential ceiling effects.

Applied Implications

Findings from the current investigation have several implications for the treatment of psychological difficulties. On a foundational level, the obtained empirical demonstration that coherence is a well established operant repertoire and that coherence contexts are generally preferred lends support to the RFT account of language and cognition that underlies the psychological flexibility model and acceptance and commitment therapy (ACT). More directly, several findings from the current investigation suggest basic mechanisms that may contribute both the maintenance and alleviation of psychological distress.

The observed finding that a small subset of participants persisted in their preference towards coherent contexts in the face of an increasingly aversive response cost may provide a basic analogue of problematic rule following within the psychological flexibility model. While the performance of these individuals was not predicted by measures of psychological inflexibility and cognitive fusion in the current study, the presence of this observed response pattern is still suggestive of costly rule following that was insensitive to direct contingencies. In particular, these participants may have been avoiding the aversive properties of ambiguity by persisting in preference towards the coherent context. This observed response pattern is consistent with the ACT conceptual account of experiential avoidance (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996), in that participants may have been responding to avoid ambiguity. While the coherence context allowed for immediate avoidance of ambiguity, it did so at the cost of additional time in the experiment, suggesting that there was a long-term negative consequence associated with the response strategy. The relationship between experiential avoidance and negative outcomes has been well established in the ACT literature (Kashdann, Barrios, Forsyth, & Steger, 2006; Ruiz, 2010).

135

The finding that ambiguity was aversive (i.e., frustrating) for participants in the study, and even more so for participants who endorsed greater levels of cognitive fusion and psychological inflexibility, also has clinical applications. This finding is congruent with the larger psychopathology literature, which suggests that intolerance of uncertainty is a maintaining feature in disorders such as Generalized Anxiety Disorder (Buhr & Dugas, 2002) and Obsessive Compulsive Disorder (Tolin, Abramowitz, Brididi, & Foa, 2003). Ambiguity is not limited to just future events, which are fundamentally unknowable, but also to past events and the reasons for why they might have occurred. Rumination can be seen as an attempt to terminate an aversive state of ambiguity (i.e., trying to figure out why something bad has happened), and it has been robustly associated with increased distress and treatment resistance in Major Depressive Disorder (Larsen, & Cowan, 1988; Lyubomirsky & Nolen-Hoeskema, 1995; Papageorgiou & Wells, 2003; Watkins & Moulds, 2005).

The existing literature suggest that sense-making, in the form of either rumination or worry, may be an effective strategy for escaping the immediate aversive state of ambiguity but at the cost of greater long term psychological distress and detriment to quality of life. Sense making under such states of aversive control, like other instances of experiential avoidance, is likely a narrow and ridged repertoire that is relatively insensitive to other consequences and contingences (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996). Interventions designed to foster increased flexibility in the presence of aversive stimulus control, such as the wide array of treatment techniques offered in the psychological flexibility model (Hayes, Strosahal, & Wilson, 2011), might be effective in reducing costly sense making maintained by avoidance of ambiguity. In particular, exercises that deliberately expose clients to states of ambiguity such as the "To Eat or Not Eat" exercise (Wilson & DuFrene, 2008, p. 133-134) may be especially effective in

136

generating increased flexibility among clients who engage in problematic sense making. Future studies should consider evaluating brief psychological flexibility interventions among individuals who display problematic sense making to determine whether or not flexible exposure to ambiguity results in greater response flexibility in this experimental preparation.

General Discussion

Considered as a whole, findings from this investigation provide empirical support for the near ubiquity of trying to make sense of ambiguous contexts and the aversive nature of ambiguity. Within the coherence testing preparation, all but one participant emitted verbal rules that were suggestive of sense making during the ambiguous task and the majority of participants responded in ways that were internally consistent and coherent with experimenter anticipated response patterns. Within the coherence preference assessment, group level analyses revealed a preference towards coherent contexts and participant level response analyses found greater rates of coherent preference among participants who entailed the necessary relations to discriminate between coherent and incoherent contexts. Within the frustration analysis, engaging in sense making was consistently associated with decreased levels of frustration and both psychological inflexibility and cognitive fusion were associated with greater levels of frustration throughout the task.

Evidence suggestive of the universal nature of coherence as a reinforcer was not found in the current study. Setting aside the methodological distinction between establishing coherence as a preference and establishing it as a reinforcer, the obtained findings only partially support the RFT assertion that coherence functions as a powerful reinforcer for derived relation responding (Hayes, Storshal, & Wilson, 2011, p. 51-52.). However, within behavior analysis and contextual behavioral science, the strength of a particular reinforcer is not treated as an absolute; rather it is

137

contextually bound and sensitive to the particulars of a given context (Michael, 1993; Hayes, 1989). The context employed in the current preparation was deliberately sterile and devoid of features that are common in naturalistic instances of sense making behavior. That is, there were no programmed consequences for engaging in coherent relational responding nor were there any instructions or other forms of socially mediated suggestion to do so. In addition, the stimuli in the task were deliberately arbitrary and lacked any pre-existing valence or relevance to participants. Thus, the emergence of a preference towards coherent contexts among a large subset of participants in this sterile preparation is suggestive that coherence alone has reinforcing properties for many individuals.

Future investigations are needed to elucidate the obtained findings. In addition to directly establishing the reinforcing properties of coherence above and beyond preference, studies should be conducted to manipulate and establish experimental control over the relative reinforcing strength of coherence. For example, studies could employ non-arbitrary and highly valenced idiographic stimuli to test the prediction that coherence is a stronger reinforcer for deriving coherent relation networks when class members are personally meaningful instead of arbitrary.

In addition, future investigations should seek to influence the relative reinforcing strength of coherence by inducing states of deprivation and satiation. Established findings have linked states of satiation to decreased preference and states of deprivation to increased preference towards potential reinforcers (Gottschalk, Libby, & Graff, 2000). Thus, deprivation, induced by preventing internally consistent responding or by providing programmed punishment for all responses in an earlier task, should increase the relative reinforcing strength of coherence and lead to increased preference towards coherent contexts. Conversely, inducing satiation by forcing internally consistent responding or by providing programmed reinforcement for all

responses in an earlier task should decrease the relative reinforcing strength of coherence and lead to decreased preference towards coherent contexts.

While additional research is needed to fully explore and experimentally manipulate the reinforcing properties of coherence, this study provides a promising preliminary account of coherence both as an operant behavior subject to antecedent control and as a potential reinforcer. This investigation also replicates and extends the earlier work of Wray (2011; see also Wray, Dougher, Hamilton, & Guinther, 2012) by assessing preference towards coherence under equal and response cost conditions and by exploring relevant moderators of resulting preference. While it is still premature to claim that coherence itself functions as a reinforcer, the obtained findings from this investigation are suggestive of the reinforcing properties of coherence.

LIST OF REFERENCES

Almada, P., Bordieri, M. J., Wilson, K. G., Kellum, K. K. & Gregg, J. A. (2010, May). *An analog study for deriving verbally experienced contingencies with eating-relevant behavior.* Paper presented at the meeting for the Association for Behavior Analysis International, San Antonio, TX.

Barnes, D., McCullagh, P. D., & Kennan, M. (1990). Equivalence class formation in non-hearing impaired and hearing impaired children. *The Analysis of Verbal Behavior*, *8*, 19-30.

Barnes-Holmes, D., Hayes, S. C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerain account of human language and cognition* (pp. 51-71). New York: Kluwer Academic / Plenum Publishers.

Barnes-Holmes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, *59*, 61-68. doi:10.1901/jeab.1993.59-61

Baumesiter, R., F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497-529. doi:10.1037/0033-2909.117.3.497

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders.* New York: International Universities Press.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd edition). New York: Guilford.

Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, *61*, 228-235. doi:10.1080/01621459.1966.10502021

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*, *5*, 47-57. doi:10.1016/S0962-1849(96)80026-6

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

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. C., Guenole, N., . . . Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire – II: A revised measure of psychological flexibility and acceptance. *Behavior Therapy*, *42*, 676-688. doi:10.1016/j.beth.2011.03.007

Bond, F. W., Hayes, S. C., & Barnes-Holmes, D. (2006). Psychological flexibility, ACT, and organization behavior. *Journal of Organizational Behavior Management*, *26*, 25-54. doi:10.1300/J075v26n01_02

Bordieri, M., Flynn, M. K., Kellum, K. K., & Wilson, K. G. (2011, May). *Generalization of derived relations to an analogue eating task: A moderator analysis*. Paper presented at the meeting for the Association for Behavior Analysis International, Denver, CO.

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin, 50*, 169–189. doi:10.1037/h0090793

Burger, J. M., & Arkin, R. M. (1980). Prediction, control, and learned helplessness. *Journal of Personality and Social Psychology, 38*, 482-491. doi:10.1037/0022-3514.38.3.482

Campbell, L. (2010). *Test-retest reliability and further validity of the cognitive fusion questionnaire* (Doctoral Dissertation). Retrieved from: http://www.era.lib.ed.ac.uk/handle/1842/5627

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.

Cella, D. F., & Perry, S. W. (1986). Reliability and concurrent validity of three visual-analogue

    mood scales. *Psychological Reports, 59*, 827-833. doi:10.2466/pr0.1986.59.2.827

Chance, P. (1999). Thorndike's puzzle boxes and the origins of the experimental analysis of

    behavior. *Journal of the Experimental Analysis of Behavior*, *72*, 433-440.

    doi:10.1901/jeab.1999.72-433

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The M.I.T Press.

Cleary, P. D., Goldberg, I. D., Kessler, L. G., & Nycz, G. R. (1982). Screening for mental

    disorder among primary care patients: Usefulness of the general health questionnaire.

    Achieves of General Psychiatry, *39*, 837-840.

    doi:10.1001/archpsyc.1982.04290070065012

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for

    cognition. *Journal of Abnormal Psychology*, *51,* 291-294.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*, 997-1003.

Coyne, L. W., & Wilson, K. G. (2004). The role of cognitive fusion in impaired parenting: An

    RFT analysis. *International Journal of Psychology and Psychological Therapy*, *4*, 469-

    486.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of

    psychopathology. *Journal of Consulting Psychology*, *1960*, *24*, 349-354.

    doi:10.1037/h0047358

Dear, P. (2006). *The intelligibility of nature: How science makes sense of the world*. Chicago,

    IL: The University of Chicago Press.

Dempster, M. A. (2009). *Development and initial validation of a scale to measure cognitive fusion* (Doctoral Dissertation). Retrieved from:

http://www.era.lib.ed.ac.uk/handle/1842/4911

Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, *2*, 36-46. doi:10.1108/13673279810249369

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, *46*, 243-257. doi:10.1901/jeab.1986.46-243

Dixon, M. R., Small, S. L., & Rosales, R. (2007). Extended analysis of empirical citations with Skinner's Verbal Behavior: 1984-2004. *The Behavior Analyst*, *30*, 197-209.

Dougher, M. J., Auguston, 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. doi:10.1901/jeab.1994.62-331

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

Doughty, A., H., Kastner, R. M., & Bismark, B. D. (2011). Resurgence of derived stimulus relations: Replication and extension. *Behavioural Processes*, *86*, 152-155. doi: 10.1016/j.beproc.2010.08.006

Drageset, J., Espehaug, B., & Kikevold, M. (2012). The impact of depression and sense of coherence on emotional and social loneliness among nursing home residents without

cognitive impairment – a questionnaire survey. *Journal of Clinical Nursing*, *21*, 965-974. doi:10.1111/j.1365-2702.2011.03932.x

Drake, C. E., & Wilson, K. G. (2008). Instructional effects on performance in a matching to sample study. *Journal of the Experimental Analysis of Behavior*, *89*, 333-340. doi:10.1901/jeab.2008-89-333

Buhr, K., & Dugas, M. J. (2002). The intolerance of uncertainty scale: psychometric properties of the English version. *Behaviour Research and Therapy*, *40*, 931-945. doi:10.1016/S0005-7967(01)00092-4

Dymond, S., May, R. J., Munnelly, A., & Hoon, A. E. (2010). Evaluating the evidence base for relational frame theory: A citation analysis. *The Behavior Analyst*, *33*, 97-117.

Dymond, S., O'Hora, D., Whelan, R., & O'Donovan, A. (2006). Citation analysis of Skinner's *Verbal Behavior*: 1984–2004. *The Behavior Analyst*, 2006, *29*, 75–88.

Fernandez, R., & Picard, R. W. (1998). Signal processing for recognition of human frustration. *Proceeding of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, *6*, 3773-3776. doi:10.1109/ICASSP.1998.679705

Fields, L., Arbtzen, E., Nartey, R. K., & Eillfsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *97*, 163-181. doi:10.1901/jeab.2012.97-163

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Ownes, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis*, *25*, 491-498. doi:10.1901/jaba.1992.25-491

Fledderus, M., Voshaar, M. A., Ten Klooster, P. M., & Bohlmeijer, E. T. (2012). Further

    evaluation of the psychometric properties of the Acceptance and Action Questionnaire-II.

    Psychological Assessment. Advanced online publication. doi:10.1037/a0028200

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study

    emotion: The example of anxiety. *Journal of Applied Behavior Analysis*, *31*, 137-156.

    doi:10.1901/jaba.1998.31-137

Fyfe, S., Williams, C., Mason, O. J., & Pickup, G. J. (2008). Apophenia, theory of mind and

    schizotypy: Perceiving meaning and intentionality in randomness. *Cortex*, *44*, 1316-1325.

Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of

    human loss avoidance. *Journal of the Experimental Analysis of Behavior*, *31*, 53-70.

    doi:10.1901/jeab.1979.31-53

Gergen, K. J., & Gergen, M. M. (1988). Narrative and the self as relationship. In L.

    Berkowitz (Ed.), *Advances in experimental social psychology*, *Vol. 21* (pp. 17-

    56). New York: Academic Press.

Gillanders, D. T., Bolderston, H., Bond, F., Dempster, M., Campbell, L., . . . Remington, B.

    (under review). The cognitive fusion questionnaire.

Goldberg D. P. (1978). *Manual of the general health questionnaire*. Windsor, England: NFER

    Publishing Co.

Goldberg, D. P., & Blackwell, B. (1970). Psychiatric illness in general practice: A

    detailed study using a new method of case identification. *British Medical*

    Journal, *2*, 439–443.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., . . . Rutter, C. (1997). The

validity of two versions of the GHQ in the WHO study of mental illness in general health

care. *Psychological Medicine*, *27,*191-197.

Gortner, E., Rude, S. S., & Pennebaker, J. W. (2006). Benefits of expressive writing in lowering

rumination and depressive symptoms. *Behavior Therapy*, *37*, 292-303.

doi:10.1016/j.beth.2006.01.004

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories,

history, and bibliography. *Intelligence*, *24*, 13-23. doi:10.1016/S0160-2896(97)90011-8

Gottschalk, J. M., Libby, M. E., & Graff, R. B. (2000). The effects of establishing operations on

preference assessment outcome. *Journal of Applied Behavior Analysis*, *33*, 85-88.

doi:10.1901/jaba.2000.33-85

Green, G., & Saunders, D. (1998). Stimulus equivalence. In K. A. Latta & M.

Perone (Eds.), *Handbook of Research Methods in Human Operant Behavior* (pp. 229-

263)*.* New York: Plenum Press.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness

and attitude reversals on a self-transforming survey. *PloS ONE*, 7*,* e45457.

doi:10.1371/journal.pone.0045457

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations

without differential consequences. *Journal of the Experimental Analysis of Behavior*, *54,*

225-237. doi:10.1901/jeab.1990.54-225

Hayes, L. J., Thompson, S., & Hayes S. C. (1989). Stimulus equivalence and rule following.

Journal of the Experimental Analysis of Behavior, 52, 275-291.

doi:10.1901/jeab.1989.52-275

Hayes*, S. C.* (Ed.). (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control*. New York: Plenum.

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. (1997). Behavior epistemology includes nonverbal knowing. In L. P. Hayes & P. M. Ghezzi (Eds.), *Investigations in behavioral epistemology* (pp. 35-43). Reno, NV: Context Press.

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

Hayes, S. C., Browstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior*, *45*, 237-256. doi:10.1901/jeab.1986.45-237

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

Hayes, S. C., Kohlenberg, B. K., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence classes, *Journal of the Experimental Analysis of Behavior*, *56*, 119-137. doi:10.1901/jeab.1991.56-119

Hayes, S. C., Levin, M. E., Plumb-Vliardaga, J., Vilatte, J. L., & Pistorello, J. (2012). Acceptance and commitment therapy and contextual behavior science: Examining the

progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy*.
Advanced online publication. doi:10.1016/j.beth.2009.08.002

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

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

Hayes, S. C., & Wilson, K. G. (1993). Some applied implication of a contemporary behavior-analytic account of verbal events. *The Behavior Analyst*, *16*, 283-301.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, *64*, 1152-1168. doi:10.1037/0022-006X.64.6.1152

Healy, H., Barnes-Holmes, Y., Barnes-Holmes, D., Keogh, C., Luciano, C., & Wilson, K. G. (2010). An experimental test of a cognitive defusion exercise: Coping with negative and positive self-statements. *The Psychological Record*, *58*, 623-640.

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relation responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*, 207-227. doi:10.1901/jeab.2000.74-207

Hesser, H., Westin, B., Hayes, S. C., & Andersson, G. (2009). Clients' in-session acceptance and cognitive defusion behaviors in acceptance-based treatment of tinnitus distress. *Behaviour Research and Therapy*, *47*, 523-528. doi:10.1016/j.brat.2009.02.002

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus

equivalence or consistent nonequivalence. *The Psychological Record*, *48*, 81-110.

Hoon, A., Dymond, S., Jackson, J. W., & Dixon, M. R. (2008). Contextual control of slot-

machine gambling: Replication and extension. *Journal of Applied Behavior Analysis*, *41*,

467-470. doi:10.1901/jaba.2008.41-467

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior.

Journal of the Experimental Analysis of Behavior, *65*, 185-242.

doi:10.1901/jeab.1996.65-185

Jackson, C. (2007). The general health questionnaire. *Occupational Medicine*, *57*, 79.

doi:10.1093/occmed/kql169

Jannoff-Bulman, R., & Frieze, L. (1983). A theoretical perspective for understanding

reactions to victimization. *Journal of Social Issues*, *39,* 1-17. doi:10.1111/j.1540-

4560.1983.tb00138.x

James, W. (1907) *Pragmatism: A new name for some old ways of thinking.* New York: Longman,

Green, and Co.

Jensen, E. (2009). Sensemaking in military planning: A methodological study of command teams.

Cognition, Technology and Work, *11*, 103-118. doi:10.1007/s10111-007-0084-x

Kalliath, T. J., O'Driscoll, M. P., & Brough, P. (2004). A confirmatory factor

analysis of the General Health Questionnaire-12. *Stress and Health*, *20,*

11-20. doi:10.1002/smi.993

Kanter, J. R. (1926). *Principles of psychology* (Vol. II). New York: Alfred Knopf.

Kanter, J. R. (1929). Language as behavior and as symbolism. *Journal of Philosophy*, *26*, 150-

158.

Kanter, J. W., Busch, A. M., Weeks, C. E., & Landes, S. J. (2008). The nature of clinical

    depression: Symptoms, syndromes, and behavior analysis. *The Behavior Analyst*, *31*, 1-

    21.

Kashdan, T. B., Barrios, V., Forsyth, J. P., & Steger, M. F. (2006). Experiential avoidance as a

    generalized psychological vulnerability: Comparisons with coping and emotion

    regulation strategies. *Behaviour Research and Therapy*, *44*, 1301-1320.

    doi:10.1016/j.brat.2005.10.003,

Kaufman, A., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd

    edition). New York: Wiley.

Kazmierczak, I. E., Strelau, J., Zawadzki, B. (2012). Search for a curvilinear relationship

    between the sense of coherence and the intensity of PTSD in MVA survivors. *Anxiety,*

    Stress & Coping: An International Journal, *25*, 443-455.

    doi:10.1080/10615806.2011.618224

Kellum, K. K. (2009). Effects of instructed, shaped, and derived methods of contacting rules on

    the fluency of rule-following and accuracy of rule-stating. *Dissertation Abstracts*

    International: Section B. Sciences and Engineering, *69*(80B), 5007.

Koopman, C., Ismailji, T., Holmes, D., Classen, C. C., Palesh, O., & Wales, T. (2005). The

    effects of expressive writing on pain, depression and posttraumatic stress disorder

    symptoms in survivors of intimate partner violence. *Journal of Health Psychology*, *10*,

    211-221. doi:10.1177/1359105305049769

Kreindler, D., Levitt, A., Woolridge, N., & Lumsden, C. J. (2003). Portable mood mapping: The

    validity and reliability of analog scale displays for mood assessment via hand-held

    computer. *Psychiatry Research*, *120*, 165-177. doi: 10.1016/S0165-1781(03)00196-3

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Larsen, R. J., & Cowan, G. S. (1988). Internal focus of attention and depression: A study of daily experience. *Motivation and Emotion*, *12*, 237-249. doi:10.1007/BF00993113

Levin, M. E., Hildebrandt, M. J., Lilis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility mode: A meta-analysis of laboratory-based component studies. *Behavior Therapy.* Advanced online publication. doi: 10.1016/j.beth.2012.05.003

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology*, *56*, 201-239.

Lips-Wiersma, M. (2002). The influence of spiritual "meaning-making" on career behavior. *Journal of Management Development*, *21*, 497-520. doi:10.1108/02621710210434638

Luciano C., Gomez-Becerra, I., & Rodriguez-Valverde, M. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior*, *87*, 349-365. doi:10.1901/jeab.2007.08-06

Lyubomirsky, S., & Nolen-Hoesksema, S. (1995). Effects of self-focused rumination on negative thinking and interpersonal problem solving. *Journal of Personality and Social Psychology*, *69*, 176-190. doi:10.1037/0022-3514.69.1.176

Maier, S. F., & Seligman, M. E. (1976). Learned helplessness: Theory and evidence. *Journal of Experimental Psychology: General*, *105,* 3-46. doi:10.1037/0096-3445.105.1.3

Maher, B. A. (1978). A reader's, writer's, and reviewer's guide to assessing research reports in clinical psychology. *Journal of Consulting and Clinical Psychology*, *46*, 835-838. doi:10.1037/0022-006X.46.4.835

Makowska, Z., Merecz, D., Mosacicka, A., & Kolasa, W. (2002). The validity of general health questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *International Journal of Occupational Medicine and Environmental Health 15*, 353-362.

Maslach, C., Schaufeil, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*, 397-422. doi:10.1146/annurev.psych.52.1.397

Masuda, A., Hayes, S. C., Sackett, C. F., & Twohig, M. P. (2004). Cognitive defusion and self-relevant negative thoughts: Examining the impact of a ninety year old technique. *Behaviour Research and Therapy*, *42*, 477-485. doi:10.1016/j.brat.2003.10.008

Masuda, A., Hayes, S. C., Twohig, M. P., Drossel, C., Lillis, J., & Washio, Y. (2009). A parametric study of cognitive defusion and the believability and discomfort of negative self-relevant thoughts. *Behavior Modification*, *33*, 250-262. doi:10.1177/014544550832625

Masuda, A., Twohig, M. P., Stormo, A., Feinstein, A. B., Chou, Y., & Wendell, W. W. (2010). The effects of cognitive defusion and thought distraction on emotional discomfort and believability of negative self-referential thoughts. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*, 11-17. doi:10.1016/j.jbtep.2009.08.006

Mendola, R. Tennen, H., Affleck, G., McCann, & Fitzgerald, T. (1990). Appraisal and adaptation among women with impaired fertility. *Cognitive Therapy and Research*, *14*, 79-93. doi:10.1007/BF01173526

Michael, J. (1993). Establishing operations. *The Behavior Analyst*, *16*, 191-206.

Mineka, S., & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annual Review of Psychology*, *1985*, 459-529. doi:10.1146/annurev.ps.36.020185.002431

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and

    publication bias. *Behavioral Ecology*, *15*, 1044-1045. doi:10.1093/beheco/arh107

Nastally, B. L., Dixon, M. R., & Jackson, J. W. (2010). Manipulation of slot machine preference

    in problem gamblers through contextual control. *Journal of Applied Behavior Analysis*,

    43, 125-129. doi:10.1901/jaba.2010.43-125

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European*

    Journal of Social Psychology, 15, 263-280.

Neshat-Doost, H. T., Dalgleish, T., Yule, W., Kalantari, M., Ahmadi, S. J., . . . Jobson, L. (2012).

    Enhancing autobiographical memory specificity through cognitive training: An

    intervention for depression translated from basic science. *Clinical Psychological Science*.

    Advanced online publication. doi:10.1177/2167702612454613

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic

    stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of*

    Personality and Social Psychology, *61*, 115-121. doi:10.1037/0022-3514.61.1.115

O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relation responding and human

    language: Evidence from the WAIS-III. *The Psychological Record, 55*, 155-176.

Ost, L. G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and

    meta-analysis. *Behavoiur Research and Therapy*, *46*, 296-321.

    doi:10.1016/j.brat.2007.12.005

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B. A., & Page, T. J. (1985). Assessment of

    stimulus preference and reinforcer values with profoundly retarded individuals. *Journal*

    of Applied Behavior Analysis, *18*, 249-255. doi:10.1901/jaba.1985.18-249

Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 429-452.

Papageorgiou, C., & Wells, A. (2003). An empirical test of a clinical metacognitive model of rumination and depression. *Cognitive Therapy and Research*, *27*, 261-273. doi:10.1023/A:1023962332399

Park, C. L., Riley, K. E., & Snyder, L. B. (2012). Meaning making coping, making sense, and post-traumatic growth following the 9/11 terrorist attacks. *The Journal of Positive Psychology*, *7*, 198-207. doi:10.1080/17439760.2012.671347

Pennebaker, J. W., & Chung, C. K. (2011). Expressive writing and its links to mental and physical health. In H. S. Friedman (Ed.), *Oxford handbook of health psychology* (pp. 417-437)*.* New York: Oxford University Press.

Pennebacker, J. W., & Seagal, J. D. (1999). Forming a story: the health benefits of narrative. *Journal of Clinical Psychology*, *55*, 1243-1254. doi:10.1002/(SICI)1097-4679(199910)55:10<1243::AID-JCLP6>3.0.CO;2-N

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

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, *316*, 1236-1238. doi:10.1136/bmj.316.7139.1236

Peterson, C., Bettes, B. A., & Seligman, M. E. P. (1982). *Spontaneous attributions and depressive symptoms.* Unpublished manuscript, Virginia Polytechnic Institute and State University.

Peterson, C., & Seligman, M. E. P. (1984). Causal explanations as a risk factor for depression: Theory and evidence. *Psychological Review, 91*, 347-374. doi:10.1037/0033-295X.91.3.347

Pierce, C. A., Block, R., A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Education and Psychological Measurement, 64*, 916-924. doi:10.1177/0013164404264848

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63*, 239-254. doi:10.1901/jeab.1995.63-239

Pilgrim, C. & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63,* 225-238. doi:10.1901/jeab.1995.63-225

Pirolli, P., & Russell, D. M. (2011). Introduction to this special issue on sensemaking. *Human-Computer Interaction, 26*, 1-8. doi: 10.1080/07370024.2011.556557

Popper, K. (2002). *The logic of scientific discovery*. New York: Rutledge Classics.

Powers, M. B., Zum Vorde Sive Vording, M. B., & Emmelkamp, M. P. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherpay and Psychosomatics, 78*, 73-78. doi:10.1159/000190790

Price, D. D., McGarth, P. A., Rafili, A., & Buckingham, B. (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain, 17*, 45-56. doi:10.1016/0304-3959(83)90126-4

Putnam, H. (1975). *Matter, Mathematics and method: Philosophical papers, volume 1*. Cambridge: Cambridge University Press.

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. (2007). Assessing relational

learning deficits in perspective-taking in children with high-functioning autism spectrum

disorder. *The Psychological Record*, *57*, 23-46.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne

Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119-125.

doi:10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Riley, J. C. (2005). Estimates of regional and global life expectancy: 1800-2001. *Population and

Development Review*, *31*, 537-543. doi: 10.1111/j.1728-4457.2005.00083.x

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief

preference assessment. *Journal of Applied Behavior Analysis*, *31*, 605-620.

doi:10.1901/jaba.1998.31-605

Roberts, G. A. (2000). Narrative and severe mental illness: What place to stories have in an

evidence-based world? *Advances in Psychiatric Treatment, 6,* 432-441.

doi:10.1192/apt.6.6.432

Roe, D., & Davidson, L. (2005). Self and narrative in schizophrenia: Time to author a new story.

Medical Humanities, *31*, 89-93. doi:10.1136/jmh.2005.000214

Ruiz, F. (2010). A review of acceptance and commitment therapy (ACT) empirical evidence:

Correlation, experimental psychopathology, component and outcome studies.

International Journal of Psychology and Consultation, *10*, 125-162.

Ruiz, F. (2012). Acceptance and commitment therapy versus traditional cognitive behavioral

therapy: A systematic review and meta-analysis of current empirical evidence.

International Journal of Psychology & Psychological Therapy, *12*, 333-357.

Samuels, R. (2004). Innateness in cognitive science. *TRENDS in Cognitive Sciences*, *8*, 136-141. doi:10.1016/j.tics.2004.01.010

Sanford, L. D., Yang, L., Wellman, L. L., Lui, X., & Tang, X. (2010). Differential effects of controllable and uncontrollable footshock stress on sleep in mice. *Sleep*, *33*, 621-630.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Beh*avior, *50*, 145–162. doi:10.1901/jeab.1988.50-145

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *49*, 95–115. doi:10.1901/jeab.1988.49-95

Shimoff, A., Catania, C., & Matthews, B. A. (1981). Unstructured human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior*, *36*, 207-220. doi:10.1901/jeab.1981.36-207

Shimoff, E., Matthew, B. A., & Catania, A. (1986). Human operant performance: Sensitivity and pseudosensitivity to contingencies. *Journal of the Experimental Analysis of Behavior*, *46*, 149-157. doi:10.1901/jeab.1986.46-149

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *1971*, *14*, 5-13.

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

Skinner, B. F. (1936). The verbal summator and a method for the study of latent speech. *The Journal of Psychology: Interdisciplinary and Applied*, *2*, 1936. doi: 10.1080/00223980.1936.9917445

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, *52,* 270-277. doi:10.1037/h0062535

Skinner, B. F. (1953). *Science and human behavior.* New York: The Free Press.

Skinner, B. F. (1957). *Verbal behavior*. Cambridge, MA: B. F. Skinner Foundation.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.) *Problem-solving: Research, method, and theory* (pp. 225-257). New York: Wiley.

Skinner, B. F. (1974). *About behaviorism*. New York: Appleton-Century.

Smith, J. M., & Alloy, L. B. (2009). A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. *Clinical Psychology Review*, *29*, 116-128. doi:10.1016/j.cpr.2008.10.003

Smyth, J. M., Hockermeyer, J. R., & Tulloch, H. (2008). Expressive writing and post-traumatic stress disorder: Effects on trauma symptoms, mood states, and cortisol reactivity. *British Journal of Health Psychology*, *13*, 85-93. doi:10.1348/135910707X250866

Sosnowski, T. (1983). Sources uncertainty and psychophysiological state: Changes under conditions of noxious stimulus anticipation. *Biological Psychology*, *17*, 297-309. doi:10.1016/0301-0511(83)90005-4

Sosnowski, T. (1988). Patterns of skin conductance and heart rate changes under anticipatory stress conditions. *Journal of Psychophysiology*, *2*, 231-238.

Spera, S. P., Buhrfeind, E. D., & Pennebaker, J. W. (1994). Expressive writing and coping with job loss. *Academy of Management Journal*, *37*, 722-733. doi:10.2307/256708

Spradlin, J. E., Saunders, K.J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29-42). Reno, NV: Context Press.

Steel, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519-555. doi:10.1901/jeab.1991.56-519

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. (5th ed.). Boston, MA: Pearson.

Tarbox, J., Zuckerman, C. K., Bishop, M. R., Olive, M. L., & O'Hora, D. P. (2011). Rule-governed behavior: Teaching a preliminary repertoire of rule-following to children with autism. *The Analysis of Verbal Behavior*, *27*, 125-139.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement 2* (4, Whole No. 8).

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., & Foa, E. B. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *17*, 223-242. doi:10.1016/S0887-6185(02)00182-2

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

Vonnegut, K. (1963). *Cat's cradle*. New York: Holt Rinehart & Winston.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work

and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics

Society*, *50*, 433-441. doi:10.1518/001872008X31215

Watkins, E. (2004). Adaptive and maladaptive ruminative self-focus during emotional

processing. *Behaviour Research and Therapy*, *42*, 1037-1052.

doi:10.1016/j.brat.2004.01.009

Watkins, E., & Moulds, M. (2005). Distinct modes of ruminative self-focus: Impact of abstract

versus concrete rumination on problem solving in depression. *Emotion*, *5*, 319-328.

doi:10.1037/1528-3542.5.3.319

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of

sensemaking. *Organization Science*, *16*, 409-421. doi:10.1287/orsc.1050.0133

Weil, T. M., Hayes, S. C., & Capurro, P. (2011). Establishing a deictic relational repertoire in

young children. *The Psychological Record*, *61*, 371-390.

Weiner, I. B. (1994). The Rorschach inkblot method (RIM) is not a test: Implications for

theory and practice. *Journal of Personality Assessment*, *62*, 1994.

doi:10.1207/s15327752jpa6203_9

Wechsler, D. (2008). *WAIS-IV: Administration and scoring manual*. San Antonio, TX:

Pearson.

Williams, J. M., Barnhofer, T., Crane, C., Herman, D., Raes, F., . . . Dalgleish, T. (2007).

Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*,

133, 122-148. doi:10.1037/0033-2909.133.1.122

Williams, J. M., Tesdale, J. D., Segal, Z. V., & Soulsby, J. (2000). Mindfulness-based cognitive

   therapy reduces overgeneral autobiographical memory in formerly depressed patients.

   Journal of Abnormal Psychology, *109*, 150-155. doi:10.1037/0021-843X.109.1.150

Wilkinson, L., & APA Task Force on Statistical Inference (1999). Statistical methods in

   psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.

   doi:10.1037/0003-066X.54.8.594

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology.

   Quarterly Review of Biology, *82*, 327-348. doi:10.1086/522809

Wilson, K. G. (2001). Some notes on theoretical constructs: Types and validation from a

   contextual-behavioral perspective. *International Journal of Psychology and*

   Psychological Therapy, 1, 205-215.

Wilson, K. G. & DuFrene, T. (2008). *Mindfulness for two: An acceptance and commitment*

   therapy approach to mindfulness in psychotherapy. Oakland, CA: New Harbinger.

Wilson, K. G., & Hayes, S. C. (1996). Resurgence of derived stimulus relations. *Journal of the*

   Experimental Analysis of Behavior, *66*, 267-281.doi:10.1901/jeab.1996.66-267

Wilson, K. G., Hayes, S. C., Gregg, J., & Zettle, R. D. (2001). Psychopathology and

   psychotherapy. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame*

   theory: A post-Skinnerain account of human language and cognition (pp. 211-237). New

   York: Kluwer Academic / Plenum Publishers.

Wilson, K. G., Whiteman, K., & Bordieri, M. J. (in press). The pragmatic truth criterion and

   values in contextual behavioral science. In S. Dymond & B. Roche (Eds.). *Advances in*

   Relational Frame Theory & Contextual Behavioral Science: Research & Application.

   Oakland, CA: New Harbinger.

Winston, A. S., & Baker, J. E. (1985). Behavior analytic studies of creativity: A critical review. *The Behavior Analyst*, *8*, 191-205.

Wray, A. (2011). *The reinforcing effects of making sense: Positive, negative or both* (Doctoral dissertation). Retrieved from http://hdl.handle.net/1928/13181

Wray, A., Dougher, M. J., & Bullard, L. (2008, May). *Examining the reinforcing effects of making sense*. Poster session presented at the annual meeting of the Association for Behavior Analysis International, Chicago, IL.

Wray, A. M., Dougher, M. J., Hamilton, D. A., & Guinther, P. M. (2012). Examining the reinforcing properties of making sense: A preliminary investigation. *The Psychological Record*, *62*, 599-622.

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. doi:10.1901/jeab.1988.50-125

Zettle, R. D., Rains, J. C., & Hayes, S. C. (2011). Processes of change in acceptance and commitment therapy and cognitive therapy for depression. A mediation reanalysis of Zettle and Rains. *Behavior Modification*, *35*, 265-283. doi:10.1177/0145445511398344

Zlomke, K. R., & Dixon, M. R. (2006). Modification of slot-machine preference through the use of a conditional discrimination paradigm. *Journal of Applied Behavior Analysis*, *39*, 351-361. doi:10.1901/jaba.2006.109-04

VITA

Michael Bordieri
mike.bordieri@gmail.com

Education & Professional Certifications

University of Mississippi   August 2009 to present
Clinical Psychology
Honors: Graduate School Fellowship (2009 to 2013), John and Lillian Wolfe Graduate Student
Achievement Award (2013), University of Mississippi Liberal Arts Graduate Student
Achievement Award (2013)

Board Certified Behavior Analyst (BCBA)   May 2010 to present
Certificate: 1-10-696

Masters of Science   August 2009
Southern Illinois University Carbondale
Behavior Analysis and Therapy
Thesis: Generating Sustainable Weight Loss: Investigating the Efficacy of a Behavior Based
Weight Loss Intervention

Bachelors of Science, Magna Cum Laude   May 2005
University of Illinois, Urbana-Champaign
Major: Psychology, Minor: History
Honors: Edmund J. James Scholar, FMC Award of Excellence Scholarship, State of Illinois
Scholar

Clinical and Research Experience

Geriatric Mental Health Therapist   October 2011 to July 2012
Region IV Community Mental Health Center – Holly Springs & Oxford, Mississippi
Delivered individual therapy and collaborated with multi-disciplinary team to provide
comprehensive mental health services to older adults in medical care facilities.

Behavior Analyst Supervisor (BCBA)   August 2011 to present
Private Practice – Oxford, Mississippi
Provided individual behavior analysis supervision to Board Certified Associate Behavior
Analysts (BCaBAs) and candidates seeking the Board Certified Behavior Analyst (BCBA)

credential. Supervision activities included assisting supervisees in case conceptualizations and treatment planning, providing consultations on ethical and professional conduct, and assisting supervisees in evaluating the efficacy of behavioral interventions.

Computer Programmer   July 2011 to present
Louisiana Board of Regents Grant – Lafayette, LA
Programmed the Body Image Flexibility Assessment Procedure (BIFAP) for Dr. Emily Sandoz using the .NET framework, designed an Access database for data management, and developed user guides and other support materials for the computer program.

Graduate Student Therapist   August 2010 to present
Psychological Services Center – University of Mississippi
Delivered individual therapy to adolescents and adults with a focus on traditional behavior therapy, contemporary behavior therapies (e.g., Acceptance and Commitment Therapy), and CBT treatment packages. Supervisors: Dr. Kelly G. Wilson, Dr. Todd Smitherman, and Dr. Alan Gross.

Mental Health Therapist   July 2010 to June 2011
Communicare (Region II Community Mental Health Center) – Pittsboro, Mississippi
Delivered individual and group therapy, provided crisis services, conducted pre-commitment evaluations, wrote integrated psychosocial intake reports, and conducted psychological consultations at a hospital emergency room.

Research Assistant   August 2009 to present
Mississippi Center for Contextual Psychology – University of Mississippi
Designed and conducted research projects, analyzed data sets, prepared research and conceptual articles for publication, provided behavior analytic and program evaluation consultation services at various agencies, and mentored undergraduate students.

Clinic Coordinator & Protocol Therapist   June 2008 to August 2009
Project HEALTH – Southern Illinois University Carbondale
Delivered Acceptance and Commitment therapy to overweight and obese adults, designed weight-loss specific therapeutic protocols, and coordinated the project's research, treatment, and training activities.

Behavior Analyst Intern   August 2008 to August 2009
Illinois Center for Autism – Fairview Heights, Illinois
Designed a computerized data collection system, delivered staff and parent training on autism and applied behavior analysis, conducted assessments, and developed behavior plans for children and adolescents with autism spectrum disorders.

Behavior Analyst Intern & Classroom Instructor   August 2007 to August 2008
Center for Comprehensive Services – Carbondale, Illinois
Wrote behavior plans, led anger management and social skills groups, prepared reports, and collaborated with a multidisciplinary team to provide rehabilitative services for adolescents with acquired brain injuries. In addition, I prepared and implemented a classroom management system, designed academic and functional lesson plans, and delivered direct instruction to adolescents with acquired brain injuries.

Teaching Experience

Undergraduate Courses
Instructor for General Psychology   Fall 2011, Spring 2012, Fall 2012, Spring 2013
Teaching Assistant for Learning   Spring 2010
Teaching Assistant for Abnormal Psychology   Fall 2009

Graduate Courses
Teaching Assistant for Graduate Seminar in ACT   Spring 2013
Teaching Assistant for Conditioning and Learning   Fall 2011
Teaching Assistant for Complex Behavior Analysis   Spring 2009

Professional Activities

Student Representative   2012-2013
Clinical Program Faculty Meetings, University of Mississippi

Program Committee Member   2012
Association for Contextual Behavior Science, World Conference X

Student Program Representative   2011-2012
Association for Behavior Analysis International

Teacher of the Year Committee (Chair)   2009
Rehabilitation Institute, Southern Illinois University Carbondale

Selected Publications

Bethay, J. S., Wilson, K. G., Schnetzer, L. W., Nassar, S. L., & Bordieri, M. J. (in press). A controlled pilot evaluation of acceptance and commitment training for intellectual disability staff. *Mindfulness*. Advanced online publication. doi: 10.1007/s12671-012-0103-8

Wilson, K. G., Whiteman, K., & Bordieri, M. J. (in press). The pragmatic truth criterion and values in contextual behavioral science. In S. Dymond & B. Roche (Eds.). *Advances in Relational Frame Theory & Contextual Behavioral Science: Research & Application*. Oakland, CA: New Harbinger.

Wilson, K. G., Bordieri, M. J., Whiteman, K. (2012). The self and mindfulness. In L. McHugh & I. Stewart (Eds.) *The Self and Perspective-Taking: Contributions and Applications from Modern Behavioral Science* (pp. 181-197). Oakland, CA: New Harbinger.

Wilson, K. G., Flynn, M. K., Nassar, S., Lucas, N., Bordieri, M. J., & Whiteman, K. C. (2012). Acceptance in cognitive behavior therapy. In W. O'Donohue & J. Fisher (Eds.) *Cognitive Behavior Therapy: Core Principles for Practice* (pp 377-398). Hoboken, NJ: Wiley.

Wilson, K. G., Bordieri, M. J., Flynn, M. K., Lucas, N., & Slater, R. (2010). Understanding acceptance and commitment therapy in context: A history of similarities and differences. In J. Herbert & E. Forman (Eds.) *Acceptance and Mindfulness in Cognitive Behavior Therapy* (pp. 233-263). Hoboken, NJ: Wiley.

Bordieri, M. J., Bordieri, J. E., & Dixon, M. R. (2008). Video golf and gambling: The impact of monetary wagers on performance. *The Analysis of Gambling Behavior*, *2*, 149-155.