6-23-2011 12:00 AM

Carving Cognition at its Joints: Insights from the Interaction between Explicit and Implicit Social Cognition

Kurt R. Peters, *The University of Western Ontario*

Supervisor: Bertram Gawronski, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

CARVING COGNITION AT ITS JOINTS:
INSIGHTS FROM THE INTERACTION BETWEEN EXPLICIT AND IMPLICIT SOCIAL COGNITION

(Spine title: CARVING COGNITION AT ITS JOINTS)

(Thesis format: Integrated Article)

by

Kurt R. Peters

Graduate Program in Psychology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. Bertram Gawronski	Dr. Joseph Cesario
Supervisory Committee	Dr. Paul Minda
Dr. James Olson	Dr. James Olson
Dr. Kali Trzesniewski	Dr. Christopher Viger

The thesis by

Kurt Ryan Peters

entitled:

**Carving Cognition at its Joints:
Insights from the Interaction between Explicit and Implicit Social Cognition**

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date

Chair of the Thesis Examination Board

Abstract

The distinction of cognition into kinds of cognitive process has proven theoretically fruitful and empirically compelling, but there remain significant challenges in deciding how best to carve cognition. First, it is unclear how to design measurement procedures that select distinct kinds of cognitive processing as exclusively as possible and, conversely, how to interpret the results of different kinds of measurement procedure. Second, the distinction between kinds of cognition must be specified with enough precision to derive empirically testable and falsifiable predictions. Third, there must be a reasonable explanation, ultimately compatible with phylogenetic evidence, for the existence of the specified distinction between kinds of cognition. The present research investigates the mutual influences between implicit and explicit self-knowledge and the influence of perceived validity on implicit and explicit evaluations. The findings challenge existing specifications of the distinction between kinds of cognition, which suggest that implicit cognition should be less sensitive than explicit cognition to situational context. As an alternative, it is suggested that the key distinction between kinds of cognition involves the capacity for *quantification*, which is a result of differences in the principles of lower-level and higher-level mental representation. Specifically, lower-level cognition is assumed to be holistic, rooted in distributed representations, whereas higher-level cognition is assumed to be symbolic, rooted in localist representations. Interaction between these processes therefore involves quantifying across holistic tokens to produce symbolic types. This perspective has important implications for theory and measurement in empirical psychology.

Keywords

Dual-Process Theories; Implicit Social Cognition; Automatic Processes; Control; Self-Concept; Attitudes

Co-Authorship Statement

The material contained in this thesis has been obtained through collaboration with Dr. Bertram Gawronski. Chapters 2 and 3 have both been published elsewhere. The written material in this thesis is my own work; however, Dr. Gawronski provided assistance with regard to the revision of the experimental papers contained in this thesis. As my thesis advisor, Dr. Gawronski also provided additional assistance in the revision of the general introduction and discussion.

Acknowledgments

First and foremost I would like to thank my advisor, Bertram Gawronski, for his help completing the research for this dissertation and for his constant support and guidance during my graduate career. His dedication to his students and to his work is truly exceptional. I would especially like to thank Bertram for allowing me the freedom to pursue my own interests in his lab, even when these sometimes conflicted with his own views. His openness to intellectual challenge reflects the best part of science, and I am sure I am a better psychologist for our many critical discussions. It has been a pleasure to work with Bertram, and I am grateful for having had the opportunity. I can only hope to emulate his example in the future.

I would also like to thank my parents for their love and support, without which I would be quite lost. I love you guys.

Thanks also to my fiancée, Kate McInnis, who has made my life so much better than it otherwise would be. Thank you for everything you've done to help me along the way, both big and small. I very likely would have ended up in the hospital with scurvy if you hadn't been here to feed me while I was writing.

I would also like to thank all my friends at Western who have made graduate school much more fun than advertised. I wish you all the best in your future endeavors.

Last but not least, I would like to thank my greyhound Rebel, who is currently sleeping upside down on the couch beside me, as he has been the majority of the time I've spent writing my dissertation. Even so, I think he still doesn't have any idea what this dissertation is about—and that helps keep it in perspective.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AMP .... Affect Misattribution Procedure

APE .... Associative-Propositional Evaluation [Model]

EPT .... Evaluative Priming Task

IAT .... Implicit Association Test

RIM .... Reflective-Impulsive Model

1 Three Problems for Carving Cognition

The desire to carve cognition into kinds is as old as psychology itself. James (1890) wrote of the distinction between an inarticulate stream of thought and a discrete train of ideas. Freud (1927) focused on the perpetual conflict between the impulsive id and the rational ego. Indeed, the creation of taxonomy—the ordering of the world into component parts and their relations—is central to science, and is especially critical during the early development of a field. Yet it is only with the recent refinement of "indirect" measurement procedures (De Houwer, 2006), which attempt to measure cognition without requiring any direct self-assessment, that the theoretically useful distinction between kinds of cognition has become empirically compelling. These procedural advances have fueled research into the relation between indirectly measured "implicit" cognition and directly measured "explicit" cognition, although the nature of this distinction, and the extent to which these measurement procedures capture it, remain challenging questions.

Within social-cognitive psychology in particular, the idea of a dual-process mind has produced a wealth of relevant data that appear to support the hypothesis that human social behavior is the product of two distinct kinds of cognition. Yet it is far from clear what this evidence tells us about the nature of these cognitive kinds—or even if carving cognition into discrete kinds obscures a more graded relation between cognitive processes. The goal of the present work is to critically examine current perspectives on the dual-process mind in social-cognitive psychology, using my own research to illuminate potential shortcomings and to suggest new theoretical directions in response.

The present chapter is structured around three key problems facing dual-process theories of cognition (Samuels, 2009). The *measurement problem* concerns how different kinds of cognition can be empirically distinguished: How well do different measurement procedures selectively assess different kinds of cognition? The *specification problem* concerns the challenge of adequately characterizing different kinds of cognition: What are the key distinctions that make them different? Finally, the *unity problem* asks a more fundamental question about the specification of these processes: Why is the mind divided into these, as opposed to some other, kinds of cognition?

Consideration of these problems reveals two general perspectives on the dual-process mind in social-cognitive psychology that differ in how they characterize higher-level, explicit and lower-level, implicit cognition. It is against this background that my own research will be presented in order to challenge assumptions underlying both of these perspectives. The final chapter will seek to make sense of these data with respect to the three problems identified above and to pursue their implications for developing a new perspective on the dual-process mind.

1.1 The measurement problem

Because cognition cannot be directly observed, psychologists must rely upon the observation of behavior to draw inferences about cognitive processes. The fact that the observation of cognition is necessarily indirect lies at the heart of the enduring problem of measurement in psychology: The link between observed behavior and underlying cognition requires an inferential leap that often seems blind. In practice, this problem makes it difficult to know what is being measured in psychological research and leaves any claims about underlying cognitive processes vulnerable to alternative explanation (Borsboom, Mellenbergh, & Van Heerden, 2004).[1] The dual-process approach to cognition provides a framework for an answer to this problem by specifying how different kinds of measurement procedure assess different kinds of cognitive process. In this way, dual-process theories can be understood as setting constraints on inferences from observed behavior to underlying cognition, providing a degree of guidance in the leap from one to the other.

---

[1] Understanding how information is mentally represented and processed is, of course, a problem unique to the cognitivist tradition of empirical psychology, which seeks to explain behavioral phenomena in terms of mediating cognitive processes. Psychologists working within the behaviorist tradition do not face this problem, since they do not appeal to cognitive processes in explaining how a stimulus causes a behavior. Given that *radical* behaviorism has failed as a basis for psychological explanation, however (e.g., Chomsky, 1959), the problem of measurement discussed here is of quite general relevance for modern empirical psychology.

From this perspective, psychological measurement procedures can be classified according to the conditions they establish for the observation of behavior. In the current state of social-cognitive psychology, indirect measurement procedures attempt to establish conditions necessary for the observation of "automatic" behavior, where automaticity is defined by the (disjunctive) presence of various functional properties of behavior (e.g., resource-efficiency, unawareness, uncontrollability, unintentionality, etc.; Bargh, 1994; De Houwer, 2006; Moors & De Houwer, 2006). These functional properties are assumed to provide the initial link between behavior and underlying cognition, such that behaviors that qualify as automatic are assumed to be produced by automatic cognitive processing (De Houwer, 2006). Hence, the functional properties that characterize a certain kind of behavior are equated to the *operating conditions* of a certain kind of cognition (Figure 1.1).

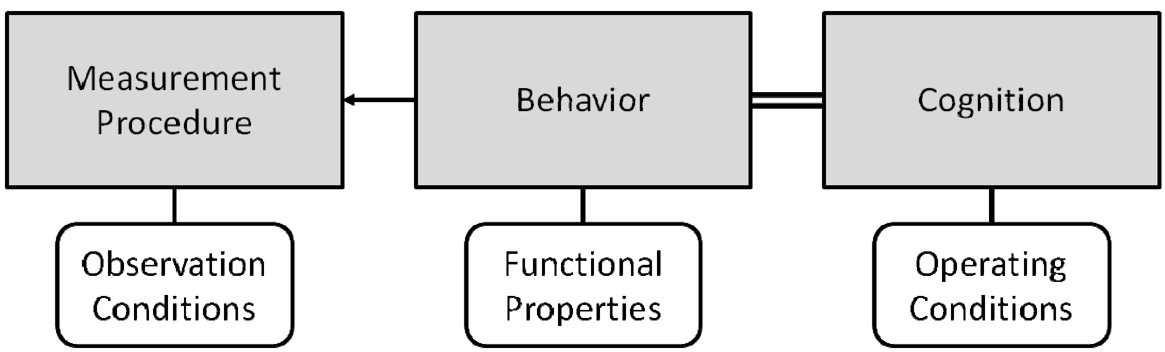

Figure 1.1. The functional properties of behavior as the mediating link in the measurement of a certain kind of cognition. In practice, the functional properties of a kind of behavior are assumed to be identical to the operating conditions of a kind of cognition (e.g., automatic behavior is produced by automatic cognition; De Houwer, 2006).

In social-cognitive research, the functional properties of automatic behavior (e.g., resource-efficiency, unawareness, unintentionality, etc.) are assumed to correspond to empirical properties of indirect measurement procedures (e.g., being multi-tasked, subliminal, speeded, etc.; De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009).

Hence, indirect measurement procedures are interpreted as revealing the operation of automatic cognitive processes. Currently, the most popular indirect measurement procedures within social-cognitive psychology are the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), the evaluative priming task (EPT; Fazio, Jackson, Dunton, & Williams, 1995), and the affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005). The IAT requires participants to quickly sort stimuli from two different categories (e.g., pictures of black vs. white persons and positive vs. negative words) into overlapping groups, and differences in reaction times are interpreted as an index of the strength of association between the categories. For example, in an IAT designed to measure automatic evaluations of black vs. white persons, the left-hand response category may be “Black person or Negative word” and the right-hand response category may be “White person or Positive word” (or the converse race/valence combinations). Usually, participants are slower to provide correct responses when the race/valence combinations are stereotypically incongruent (i.e., “Black person or Positive word”) compared to when they are stereotypically congruent (i.e., “Black person or Negative word”), suggesting an automatic evaluative bias (Nosek et al., 2007). The EPT is procedurally simpler, as it requires participants only to indicate whether a target word is positive or negative as quickly as possible; however, each word is briefly preceded by a prime (e.g., a picture of a black or white person), the valence of which is assumed to influence the reaction time of evaluative decisions about the target words. For example, in an EPT designed to measure automatic evaluations of black vs. white persons, the typical finding is that correct evaluations of negative words are faster following primes of black persons, whereas correct evaluations of positive words are faster following primes of white persons (Fazio et al., 1995). Finally, in the AMP, participants are required to indicate whether a Chinese pictograph (assumed to be meaningless) is more or less visually pleasing than average. As with the EPT, each pictograph is preceded by a prime, and the typical finding is that the valence of the prime influences ratings of the following pictograph—even when participants are explicitly admonished to avoid its influence (Payne et al., 2005).

In contrast to indirect measurement procedures, which are designed to reduce the influence of introspective self-assessment on responses, direct measurement procedures simply allow participants to report what they think. To do so, direct measurement procedures attempt to establish conditions necessary for the observation of "controlled" behavior, which is defined in contrast to the functional properties of automatic behavior (e.g., resource-inefficiency, awareness, controllability, intentionality, etc.). The functional properties of controlled behavior are assumed to correspond to empirical properties of direct measurement procedures (e.g., being non-distracted, supraliminal, non-speeded, etc.), and direct measurement procedures are therefore interpreted as revealing the operation of controlled cognitive processes (see Figure 1.1). The most common form of direct measurement procedure in social-cognitive psychology is the basic self-report item or questionnaire. For example, in investigations of racial evaluations, direct measurement procedures often rely upon self-reported evaluations of black and white persons, frequently taking the shape of a "feeling thermometer" that allows participants to rate their warmth toward a given social group. These ratings are typically interpreted in terms of deliberate (as opposed to automatic) evaluative biases, and are often found to be dissociated from indirectly measured, automatic biases (Nosek et al., 2007).

Currently, the distinction between indirect and direct measurement procedures in social-cognitive psychology is not sharp, largely because there is no standard set of observation conditions (e.g., being multi-tasked, subliminal, speeded, etc.) that a procedure must establish to qualify as one or the other. Moreover, it is unclear if any of the functional properties traditionally used to distinguish between automatic and controlled behavior (e.g., resource-efficiency, awareness, intentionality, etc.) should be regarded as necessary or sufficient (De Houwer, 2006). As a result, these properties and their empirical realizations in measurement procedures are typically treated as disjunctive, leading to measurement procedures that combine both automatic and controlled properties (e.g., speeded self-report; Ranganath, Smith, & Nosek, 2008). The use of such procedures—and the difficulty they pose for interpretation—highlight the deeper issue of how selectively any measurement procedure can assess automatic vs. controlled cognition, an issue that will be discussed in more detail shortly (see §1.1.1).

The correspondence between the observation conditions established by a certain kind of measurement procedure (e.g., indirect vs. direct) and the operating conditions of a certain kind of cognitive process (e.g., automatic vs. controlled) is a fundamental psychometric assumption in cognitivist psychology—hence I will refer to it as the *psychometric covariation thesis*. Indeed, all cognitivist psychological research is predicated on this initial link between the conditions under which behavior is observed and the conditions under which cognition is assumed to operate; when left unspecified, the default assumption is simply that there is no theoretically meaningful variation in the operating conditions of cognition, and therefore that any observation conditions established by a measurement procedure correspond to the same (and only) operating condition of cognition. With regard to the received view of dual-process cognition, however, the assumption is that the observation conditions established by a measurement procedure do distinguish, in a theoretically meaningful way, between two different operating conditions of cognition. For example, a procedure that requires speeded responses to stimuli while combined with a distracter task would generally be assumed to measure automatic behavior, and hence to reflect automatic cognitive processing, whereas a procedure that places no time or capacity constraints on responses (e.g., a standard questionnaire) would be assumed to measure controlled behavior, and hence to reflect controlled cognitive processing.

The key insight of dual-process theories of cognition lies specifically in the assumption that there is more than one *theoretically meaningful* operating condition for cognition, which is to say that different operating conditions correlate with different *kinds* of cognitive process. In particular, the distinction between kinds of cognitive process becomes theoretically meaningful when different processes are assumed to process information in different ways (e.g., via syllogistic reasoning vs. similarity; Sloman, 1996). Dual-process theories of cognition are thus defined by how they specify the covariation between the *operating conditions* of cognitive processes (via the observation conditions established by a kind of measurement procedure) and the *operating principles* that characterize different kinds of cognitive process (see Figure 1.2). I will refer to this second link, between the operating conditions and operating principles of cognition, as the *cognitive covariation thesis*.

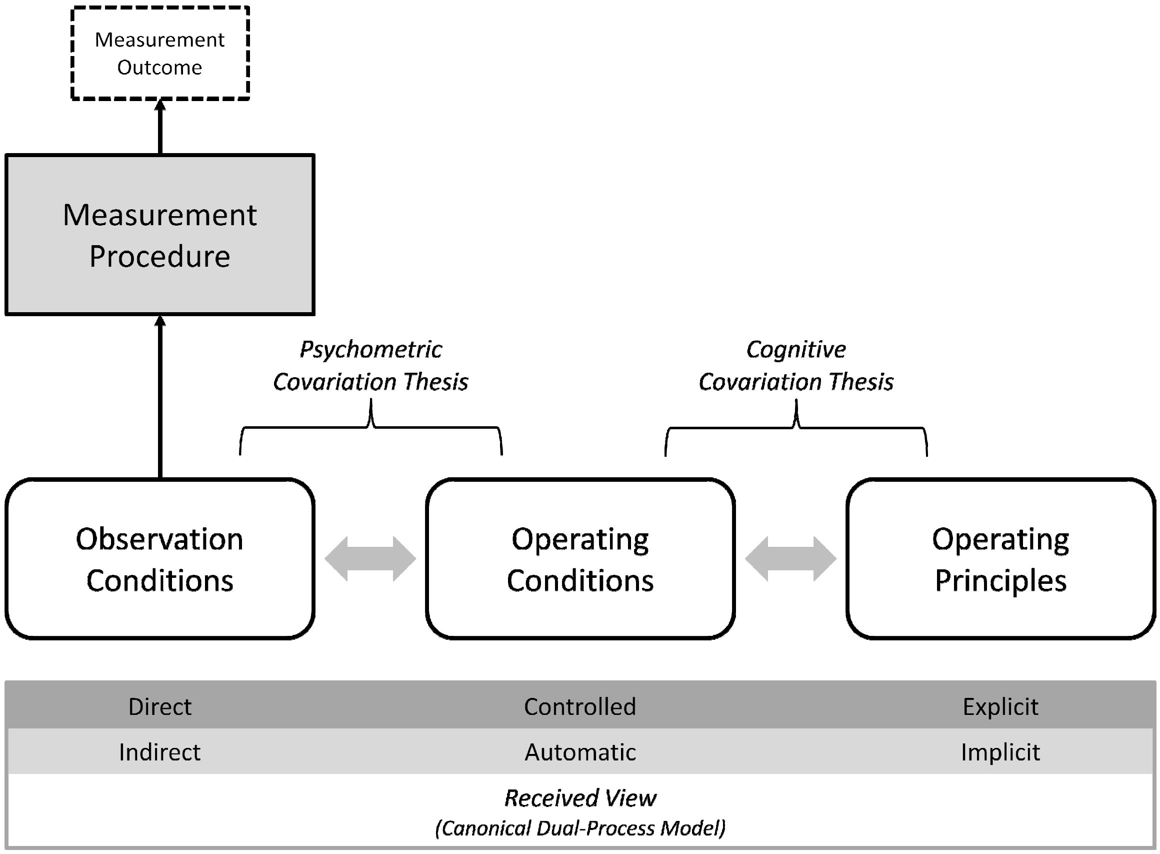

Figure 1.2. The psychometric and cognitive covariation theses as a function of the observation conditions established by a measurement procedure, the operating conditions of cognition, and the operating principles of cognition. The distinctions drawn between each of these criteria according to the "received view" of dual-process social-cognition are noted.

From this perspective, the received view of dual-process cognition (sometimes referred to as the "dual-systems" approach) is canonically defined in terms of covariation between (1) the conditions of automaticity and the principles of implicit processing and (2) the conditions of control and the principles of explicit processing. Although there are important differences between individual theories (see §1.2), the received view of dual-process cognition offers a general characterization of the difference between explicit and implicit cognition that informs much of the dual-process theorizing in social-cognitive psychology. On this view, explicit cognition operates according to the principles of syllogistic (rule-based) reasoning and is highly flexible in terms of the information it can process. Implicit cognition, in contrast, operates according to the principles of similarity (e.g., between a perceived stimulus and an object in memory), and is constrained to the associative processing of information available in the immediate situation (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Sloman, 1996; Smith & DeCoster, 2000; Strack & Deutsch, 2004).

The essential idea of *covariation*, as discussed here, encapsulates the problem of measurement for dual-process theories of cognition. The problem comes down to how strong the psychometric and cognitive covariations really are (Figure 1.2). If both covariations are strong, then the initial link between the observation conditions established by a measurement procedure and the operating conditions of cognition can be used to draw an equally strong (though necessarily indirect) inference from that measurement procedure to the kind of cognition it is measuring. This is, of course, the hope of dual-process theorists—that indirect measurement procedures allow observation of automatic, *and hence implicit*, cognition, whereas direct measurement procedures allow observation of controlled, *and hence explicit*, cognition. If, however, either covariation is weak, then the inference from observation conditions to operating principles does not go through, leaving the characterization of the cognitive processing underlying observed behavior much less constrained, and hence difficult to describe in the precise terms of a psychological mechanism.[2]

For researchers theoretically committed to the existence of multiple kinds of cognition, the covariation theses offer a natural means to resolve the problem of measurement. Yet both of these theses have also been criticized as the *cause* of this problem, for both statistical and conceptual reasons. To move research on dual-process cognition forward, it is necessary to make sense of these conflicting views. Toward that end, I will first discuss two problems endemic to dual-process measurement that have led some theorists

---

[2] The present discussion generalizes to all multi-process theories, although the focus will be on dual-process theories in particular given their dominance in the social-cognitive literature (cf. Sherman, 2006). In principle, an *n*-process model can be specified in terms of both the psychometric and covariation theses, given pairwise relations between *n* kinds of observation conditions, operating conditions, and operating principles.

to criticize the covariation theses. I will then attempt to clarify what this criticism entails and whether the theses have the potential to offer a solution to the problem of measurement.

1.1.1 Problems in the measurement of dual-process cognition

1.1.1.1 Statistical problems

A first problem in the interpretation of empirical evidence for the distinction between implicit and explicit cognition hinges on how indirect and direct measurement outcomes are statistically compared. A common approach is to draw conclusions about correspondence vs. dissociation between implicit and explicit cognition based on the zero-order correlation between indirect and direct measurement outcomes; thus, low correlations are interpreted as supporting the distinction between two kinds of cognitive process.

There are, however, a number of problems with this interpretation due to the many alternative factors that can attenuate statistical correlations (Hofmann, Gschwendner, Nosek, & Schmitt, 2005). For example, indirect measurement outcomes typically contain a greater proportion of random measurement error compared to direct measurement outcomes, which reduces the strength of correlations between the two (e.g., Bosson, Swann, & Pennebaker, 2000; Cunningham, Preacher, & Banaji, 2001). Beyond random measurement error, there may be systematic sources of error variance specific to indirect procedures, such as the presentation order of critical trials or practice effects on the IAT (Nosek, Greenwald, & Banaji, 2007), that attenuate correlations. At a more conceptual level, there is also the problem of construct correspondence between procedures, which requires that the stimuli used in parallel direct and indirect procedures be matched as closely as possible (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Payne, Burkley, & Stokes, 2008; see also Ajzen & Fishbein, 2005). These problems are inherent in dual-process research, given that such research rests upon a distinction between empirical observation conditions that opens the door to unwanted variation between different kinds of measurement procedure. Considering the very different observation conditions that typically characterize direct vs. indirect measurement procedures (e.g., speeded vs. non-speeded, subliminal vs. supraliminal, etc.), it should not be surprising that low statistical correlations between measurement outcomes can often be explained on methodological grounds.

Several responses to this problem are available, though none is perfect. One approach is to attempt to reduce measurement error by increasing the reliability (internal consistency) of indirect measurement procedures. For example, Gawronski, Cunningham, LeBel, and Deutsch (2010) found that controlling attention to specific features of a prime stimulus (i.e., race vs. age) during a Black/White EPT increased the reliability of the procedure. To reduce the influence of method-specific variance, a finding can be replicated using multiple indirect measurement procedures. The best solution, however, appears to require abandoning the correlational approach to dissociation in favor of an experimental approach, in which a single experimental manipulation is shown to produce different effects on parallel direct and indirect measurement procedures (Hofmann & Wilson, 2010). In contrast to the correlational approach, this approach compares the *same* kind of measurement procedure across experimental conditions, thereby eliminating confounds arising from comparisons between direct and indirect measurement procedures.

The experimental approach to dissociation is not a panacea, however. For one thing, the low reliability of indirect measurement procedures adversely affects the replicability of experimental findings, which can contribute to incorrect conclusions about the distinction between kinds of cognition (LeBel & Paunonen, 2011). In addition, construct correspondence remains an issue, as no matter how closely the experimenter seeks to match stimuli between direct and indirect measurement procedures, it remains unclear which features of the stimulus and context are driving responses under different operating conditions (Hofmann et al., 2005). For example, controlled responses may be inherently more (or less) context-sensitive than automatic responses, and context-sensitivity may even vary between particular indirect measurement procedures (Gawronski et al., 2010). This question bears strongly upon the specification of the cognitive covariation thesis, and will be discussed in more detail below (see §1.2).

1.1.1.2 Conceptual problems

A second problem in the interpretation of empirical evidence for the distinction between implicit and explicit cognition hinges on the notion of *process purity*, which is the assumption that measurement procedures are pure reflections of a single kind of cognitive process, uncontaminated by any other kind of processing. The assumption of process purity is widely regarded as a fallacy within social-cognitive psychology, and substantial evidence suggests that existing indirect and direct measurement procedures cannot be regarded as *selective* measures of automatic and controlled processes, or alternatively, of implicit and explicit processes (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Gawronski & Bodenhausen, in press; Sherman, Klauer, & Allen, 2010). This stance has produced a peculiar conflict for dual-process researchers: On one hand, current dual-process theories are essentially claims about the covariation between distinct *kinds* of measurement procedures, cognitive operating conditions, and cognitive operating principles; on the other hand, the assumption that this covariation is strong is criticized as leading to the misinterpretation of data.

To make sense of this conflict, it helps to examine the notion of process purity with respect to the two covariation theses separately, since each thesis independently bears upon the interpretation of data. In other words, measurement procedures might be assumed to be pure reflections of particular operating conditions (e.g., automaticity vs. control), or they may be assumed to be pure reflections of particular operating principles (e.g., implicit vs. explicit). In the former case, the process purity critique is understood as a failure of the psychometric covariation thesis: that is, as insufficient covariation between the observation conditions established by a measurement procedure and the operating conditions of a kind of cognitive process (Figure 1.2). From the canonical dual-process perspective, the failure of the psychometric covariation thesis is implied by evidence that (a) indirect measurement procedures are not pure measures of automatic cognition and that (b) direct measurement procedures are not pure measures of controlled cognition, such that behavioral responses measured using either kind of procedure appear to reflect a mixture of automatic and controlled processing. To the extent this is the case, the standard practice of using indirect and direct measurement procedures to distinguish between automatic and controlled cognition is undermined.

The strongest evidence against the process purity of measurement procedures comes from the application of mathematical modeling to indirect measurement outcomes (Sherman et al., 2010). Mathematical modeling is an analytic technique that mathematically decomposes responses on a measurement procedure into underlying processes. Applied to indirect measurement procedures, modeling techniques suggest that the cognitive processing underlying behavioral responses is not purely automatic, but is also (to a variable extent) controlled (e.g., Conrey et al., 2005; Payne, 2001). The strong evidence against the process purity of existing indirect measurement procedures suggests that mathematical modeling is indispensible for the interpretation of data testing dual-process theories. In practice, this means that measurement outcomes must be routinely decomposed into automatic and controlled components to draw inferences about dual-process cognition. Some proponents of mathematical modeling go even further, advancing an a priori argument against the psychometric covariation thesis (e.g., Jacoby, 1991; Sherman, 2008; Sherman et al., 2010). The claim is that because behavioral responses on any measurement procedure are inevitably influenced by both automatic and controlled processes, behavioral measurement procedures are *in principle* incapable of isolating a single cognitive operating condition.

Even if the psychometric covariation thesis were assumed sound, however, the process purity critique can be understood in a second, independent sense as a failure of the cognitive covariation thesis: that is, as an insufficient covariation between the operating conditions of different kinds of cognitive process and the operating principles that characterize them. In fact, Gawronski and Bodenhausen (2009) originally identified the cognitive covariation thesis (although they did not refer to it by that name) in order to criticize it on the grounds of process purity. Specifically, evidence increasingly suggests that the automatic/controlled distinction crosscuts the implicit/explicit distinction, given standard assumptions about how implicit and explicit processes operate (i.e., implicit processes being associative and explicit processes being rule-based; Sloman, 1996). For example, Gawronski, LeBel, and Peters (2007) reviewed evidence demonstrating that implicit cognition is not necessarily any less conscious, controllable, or context-sensitive than explicit cognition.

In light of this evidence, Gawronski and Bodenhausen (2009) argued that kinds of cognition should be empirically distinguished primarily with respect to their operating principles rather than with respect to the canonical dual-process operating conditions (i.e., automaticity vs. control). For example, according to Gawronski and Bodenhausen's (2006; in press) Associative-Propositional Evaluation (APE) model, the key distinction between implicit ("associative") and explicit ("propositional") cognition is the sensitivity to perceived validity, such that only propositional processes are qualified by the perceived validity of information. Thus, Gawronski and Bodenhausen (in press) claim that "there is no one-to-one mapping between operating principles and operating conditions, such that associative processes would operate automatically, whereas propositional processes operate in a controlled fashion…. Instead, both associative and propositional processes have automatic and controlled aspects" (p. 6). From the perspective of the APE model, then, Gawronski and Bodenhausen's (2009) proposal appears to require distinguishing between kinds of cognition post-hoc, based on the empirical sensitivity to perceived validity in a particular processing situation, rather than a priori, based on the functional properties of cognition.[3]

1.1.1.3 Summary

The problem of measurement in dual-process research raises the question of how different kinds of process can be measured and, conversely, how to know what kind (or kinds) of process a measurement procedure reflects. The statistical problems inherent in comparing measurement outcomes from different kinds of procedure are substantial but

---

[3] This proposal might also be interpreted as calling for a re-specification of the operating conditions of cognition in terms of *different* functional properties, which promote or suppress the influence of perceived validity (as opposed to the standard properties used to distinguish automaticity vs. control; Bargh, 1994). From this perspective, Gawronski and Bodenhausen's (2009) criticism of the cognitive covariation thesis for failing to maintain process purity is not a rejection of the thesis in principle, but is rather a call to revise its specification. It remains an empirical question whether different operating conditions can be identified that successfully discriminate between associative and propositional processing (i.e., in terms of sensitivity to perceived validity).

become more tractable in experimental research. On the other hand, the conceptual problems arising from assumptions about the process purity of these different measurement procedures appear more troublesome, calling into question the validity of both the psychometric and cognitive covariation theses. From this perspective, the problem of measurement in dual-process research appears to be *caused* by the covariation theses, as each implies an assumption of process purity (between observation conditions and operating conditions, and between operating conditions and operating principles) that can lead to the misinterpretation of data. Yet I have also suggested that these theses are vital not only to dual-process research but to the practice of cognitivist psychology in general. The next section attempts to resolve these conflicting views.

1.1.2 The value of the covariation theses

To understand the value of the covariation theses in dual-process research, it is necessary to examine both of the process-purity criticisms discussed above in more detail.

1.1.2.1 Defending the psychometric covariation thesis

First, with regard to the psychometric covariation thesis, it has been argued that behavioral responses *inevitably* reflect the operation of cognition under multiple conditions (in particular, reflecting both automatic and controlled cognition), and therefore that any behavioral measurement procedure will inevitably be process-impure (Jacoby, 1991; Sherman, 2008; Sherman et al., 2010). From this perspective, it is logically impossible for the observation conditions established by a measurement procedure to select a specific kind of cognitive operating condition—implying a rejection of the psychometric covariation thesis *in principle*.

This argument, however, begs the question of whether behavioral responses are inevitably process-impure. In fact, the process purity of behavioral responses is an empirical question, which will depend on how the distinction between kinds of process is specified. Consequently, the psychometric covariation thesis cannot be rejected on logical grounds. It is entirely reasonable to expect that this thesis would hold given a more appropriate specification—that is, by identifying the specific observation conditions and operating conditions that *do* covary. Thus, although evidence from mathematical modeling suggests that responses on current measurement procedures are the product of a combination of automatic and controlled processes, it would be incorrect to interpret this empirical fact as implying the rejection of *any* specification of the psychometric covariation thesis: Empirical evidence cannot rule out the possibility that there is a different distinction between observation conditions (other than that between direct and indirect) that would cleanly distinguish between cognitive operating conditions (in terms of automaticity vs. control or some alternative distinction; see §4.2.2.1 for such a proposal).

1.1.2.2 Defending the cognitive covariation thesis

Based on evidence that automatic and controlled conditions do not select purely implicit or explicit cognitive processing, Gawronski and Bodenhausen (2009) argued that the automatic/controlled distinction should not be used as a guide to distinguishing between the operation of implicit and explicit cognition. This negative claim does not necessitate the rejection of the cognitive covariation thesis in principle, as it leaves open the possibility that a revised specification of operating conditions and operating principles could covary more strongly. Nevertheless, Gawronski and Bodenhausen (2009, in press) appear to claim that the principle of covariation should be abandoned (though see Footnote 3 for an alternative interpretation). In practice, this means that there is no a priori way to test dissociations between implicit and explicit processes, since there is no established link between operating conditions and operating principles that can guide the choice of measurement procedure to reveal the operation of explicit vs. implicit cognition (Figure 1.2). This not only makes it difficult to formulate experimental designs to test predicted dissociations, but it also leaves multi-process theorizing disorganized, since there are no particular conditions of cognitive operation (and hence of behavior) that are characterized by specific kinds of cognitive processing (i.e., operating principles). From this perspective, researchers might claim that there are two (or more) kinds of cognition but would have no means to specify *when* one or the other would be expected to influence behavior. More fundamentally, this approach does not eliminate the necessity of the cognitive covariation thesis for psychological research. Operating principles are theoretical properties that can be empirically identified only *indirectly* via their covariation with operating conditions (and, ultimately, observation conditions; Figure 1.2). The fact that automatic processing does not appear to be purely associative, or controlled processing purely rule-based, does not imply that the assumption of covariation between operating conditions and operating principles should be abandoned, but rather that their specification may be in need of revision.

1.1.2.3 Two choices

The broader point to be made here, with respect to both covariation theses, is that they are fundamental assumptions required for empirical research in cognitivist psychology—they cannot simply be rejected, since doing so would make the inferences (from behavioral observation to cognitive process) on which cognitivist psychological explanation is based unwarranted (see Figure 1.2). Rather, it is the specific *form* of each thesis that should be subjected to criticism, not the principle of covariation itself. In the absence of a dual-process theory that explicitly specifies the form of these theses, cognitivist research is nevertheless predicated on an assumption of uniformity, such that all kinds of measurement procedure reflect the same (and only) operating principles of cognition, regardless of variability in operating conditions. Although this assumption sounds rather simple, it constitutes a theoretical claim about covariation no less significant than that of the most complex multi-process theory (see Kruglanski, Erb, Pierro, Mannetti, & Chun, 2006, for one effort to develop such a single-process model). Consequently, given the necessity of the covariation theses for cognitivist research, dual-process theorists face a choice between working with process-impure (i.e., weak) covariation theses and attempting to refine the resulting data, or striving to specify process-pure (i.e., strong) covariation theses.

A priori arguments about process purity aside, the availability of sophisticated mathematical modeling techniques for interpreting psychological data might be seen as compensating for weak covariation theses. From this perspective, there is no need to strive for process-pure theses if weaker versions can be "corrected" with the systematic use of mathematical models. The problem with this argument is that data (whether the product of behavioral observation or mathematical analysis) cannot unambiguously "correct" theory: There is a dialectical relation between the two, with data constraining theoretical claims and theory informing the interpretation of data. In particular, mathematical modeling is not theory-free, since modeling techniques themselves require theoretical assumptions in order to specify model parameters—that is, assumptions about how to divide cognition into different kinds of process and about the conditions under which each kind of process will produce a certain behavioral outcome (Payne & Bishara, 2009; Sherman et al., 2010). Thus, mathematical modeling fails to compensate for weak covariation theses because—resting on process assumptions itself—it cannot be viewed as a bedrock source of data for evaluating theoretical claims about the operation of distinct psychological processes in the interpretation of behavioral data.

The alternative is to focus on developing stronger covariation theses that approach process-purity. This choice implies a stronger, and hence more functional, dialectic between theory and data in dual-process research. Indeed, the psychometric and cognitive covariation theses jointly constitute a description of this dialectic, running from observed behavior to theoretical process and back (Figure 1.2). A dual-process researcher can therefore start from measurement procedures, try to determine what operating conditions the established observation conditions correspond to, and then try to determine the operating principles that characterize processing under those conditions; based on empirical feedback, the researcher can then revise operating principles, which should inform assumptions about the conditions under which those processes operate, and then seek to modify or design measurement procedures to correspond to those conditions. These approaches are complementary and are equally valuable. Given that the psychometric and cognitive covariation theses are indispensible for cognitivist research, then, the most practical response to the problem of process purity is to view any specification of observation conditions, operating conditions, and operating principles as *provisional* distinctions susceptible to revision, rather than viewing any particular specification as inevitable (and inevitably flawed).

1.1.3 Summary

The problem of measurement in dual-process research is reflected in conflicting views of the psychometric and cognitive covariation theses. The principle of covariation linking observed behavior to inferred process (Figure 1.2) would seem to be a key insight of dual-process theories, but it has also been criticized for encouraging assumptions about process purity that can lead to the misinterpretation of behavioral data. I have argued that the problem is not the principle of covariation itself, which is necessary for cognitivist research, but rather how to specify that covariation. Thus, although data (whether the product of behavioral observation or mathematical analysis) can speak to the adequacy of particular *specifications* of these theses, they cannot be used to reject them *in principle*. Instead, data should be used to revise the two covariation theses in order to make their specification more empirically adequate. Since dual-process research has primarily focused on how to characterize the distinction between different kinds of cognitive process in terms of their operating principles, the next section will consider the problem of how to specify the cognitive covariation thesis in particular.

1.2 The specification problem

The specification problem in dual-process research can be treated with respect to both the psychometric and cognitive covariation theses. As discussed above, the conjunction of these theses constitutes a dialectic between observable behavior and theoretical process that can be approached from either direction. Specification of the psychometric covariation thesis is undoubtedly an important question, and dual-process researchers have increasingly turned their attention to this challenge (De Houwer, 2008; De Houwer et al., 2009; Gawronski, Deutsch, LeBel, & Peters, 2008). It is the specification of the cognitive covariation thesis, however, that has been of primary interest to dual-process researchers: How is cognition carved into different kinds of process? How are these kinds characterized, and what is the key (necessary and sufficient) distinction between them? In pursuing these questions, researchers have generally assumed the standard specification of the psychometric covariation thesis, which equates indirect measurement procedures with the conditions of automaticity and direct measurement procedures with the conditions of control. Although this assumption is vulnerable to empirical process purity criticisms, it is (in some form) necessary for testing theories about the operating principles that characterize different kinds of psychological process. Moreover, empirical feedback from these tests ultimately informs the design of measurement procedures as a result of the observation/process dialectic—an implication that will be explored later in light of the present research (see §4.2).

The current section will focus on two major attempts to specify the cognitive covariation thesis in social-cognitive psychology. Many dual-process theories have been proposed to explain various aspects of social behavior, including evaluation, persuasion, and impression formation. These theories differ in detail, but they share fundamental similarities in how they characterize the distinction between automatic and controlled cognitive processes. In particular, the assumption is that, by default, automatic processing provides the input to controlled processing, which serves a regulatory or "corrective" function. Dual-process theories of social-cognition generally describe this relation in terms of distinct *lower-level*, automatic, and *higher-level*, controlled kinds of cognitive process. These attempts to specify the cognitive covariation thesis constitute hypotheses about how to carve cognition. Thus, one approach has been to distinguish between kinds of cognitive process in terms of the degree of sensitivity to the situational context, which I will refer to as the *contextualization hypothesis*. Another approach has been to distinguish between kinds of cognitive process in terms of sensitivity to perceived validity, which I will refer to as the *validation hypothesis*. Admittedly, this analysis might be criticized for painting the crowded field of dual-process theories of social-cognition with a broad brush, missing their many nuances. Yet this level of abstraction is necessary for identifying *precisely* what each model assumes to be the necessary distinction between kinds of cognition, as opposed to merely probabilistic or "symptomatic" features of their operation (Samuels, 2009). Indeed, to the extent that any given dual-process theory cannot be described in terms of at least one necessary distinction between kinds of cognition, it becomes difficult to falsify and loses empirical value.

Below, I will first outline the basic assumptions underlying the two major attempts to specify the cognitive covariation thesis. I will then identify specific models of each type that have been proposed to explain various aspects of social behavior, along with their empirical successes and challenges. Afterward, I will address the related question of the number of mental representations required by dual-process theories, which will be relevant to the interpretation of the present research.

1.2.1 The contextualization hypothesis

1.2.1.1 Basic assumptions

The philosopher John Locke proposed that complex ideas are built out of combinations of simple ideas, which are themselves irreducible. The notion that thought is *compositional* in this way underlies the contextualization hypothesis about how to carve cognition. From this perspective, the key distinction between lower-level and higher-level cognition is the level of complexity, or contextual detail, of the information upon which each kind of process operates. The basic idea is that the distinction between lower-level and higher-level cognition is marked by a bottom-up transition from general, memory-based categories to individuated instances through the higher-level integration of situational information; in other words, automatic processes operate upon types, whereas controlled processes, by integrating situational details, individuate types into tokens. For example, when applied to impression formation, the contextualization hypothesis suggests that early, automatic reactions toward a social target will reflect relatively rough categorical representations (i.e., stereotypes), whereas later, controlled reactions may be influenced by details in the immediate situation that can be used to individuate the target (e.g., Fiske & Neuberg, 1990). Because of this type-to-token dynamic, dual-process models of this sort are naturally structured by a *monitoring-and-correction* processing schema, in which higher-level, controlled processing serves to “correct” automatically activated, lower-level category knowledge as the situation demands. Consequently, these models are often assumed to operate primarily in a bottom-up fashion, such that lower-level processing influences higher-level processing, but not vice versa.

The basic distinction between type-dependent processing and token-dependent processing specified by the contextualization hypothesis assumes that behavioral responses driven by automatic processing are rooted in stable category knowledge stored in memory, and hence should be insensitive to situational context. In contrast, behavioral responses driven by controlled processing are assumed to reflect the integration of individuating contextual information, and hence should be more situationally appropriate. Typically, the type/token distinction is augmented by an additional assumption about how type-dependent and token-dependent cognitive processes operate. Specifically, automatic, type-dependent processing is assumed to be *associative*, such that it is purely a function of similarity; controlled, token-dependent processing is assumed to be *rule-based*, characterized by the principles of syllogistic reasoning (see Table 1.1). From this perspective, higher-level, explicit knowledge should be learned quickly as a function of reasoning, whereas lower-level, implicit knowledge—being immune to the top-down influence of rule-based processing—should be learned slowly as a function of repeated associations (Smith & DeCoster, 2000).

The Contextualization Hypothesis	
Operating Conditions	
Automatic	Controlled
Unintentional	Intentional
Uncontrollable	Controllable
Resource-efficient	Resource-inefficient
Outside awareness	Within awareness
Operating Principles	
Implicit	Explicit
Situation-insensitive*	Situation-sensitive*
Associative	Rule-based

Table 1.1. The specification of the cognitive covariation thesis per the contextualization hypothesis.
*Denotes the key distinction between cognitive operating principles.

An important consequence of the equation of situation-insensitive processing with associative principles, and of situation-sensitive processing with rule-based principles, is that these two kinds of cognition become imbued with significant normative connotations. Thus, similarity-based processing across rough, memory-based categories will frequently appear unintelligent, rigid, and potentially maladaptive. In contrast, rule-based processing of detailed, context-sensitive tokens will appear intelligent, flexible, and adaptive. Returning to the example of impression formation, the contextualization hypothesis suggests that automatic evaluations of a social target will reflect stable, stereotypical knowledge, which is automatically activated merely as a function of similarity between the target and a social category. Lacking the "corrective" influence of rule-based processing, which can deliberately qualify evaluations of the target based on individuating information, automatic evaluations would in many situations be expected to lead to unwarranted discrimination (e.g., Devine, 1989). Despite the fact that these normative connotations remain tacit in the social-cognitive literature (or perhaps for this reason), they have played a significant role in shaping dual-process research in the field. For example, the standard distinction between the operating conditions of automatic and controlled cognition (i.e., in terms of resource-efficiency, awareness, controllability, and intentionality) is largely informed by the conditions under which cognitively simple (unintelligent) vs. complex (intelligent) processes would be expected to operate (Greenwald, 1992). Moreover, these connotations perpetuate the folk taxonomy of cognition (i.e., rational/reflective vs. irrational/impulsive), which continues to color theorizing about the cognitive processes underlying human behavior.

1.2.1.2 Empirical applications and challenges

The contextualization hypothesis characterizes many of the classic dual-process models of social-cognition. These early models were generally inspired by the belief that organisms need to quickly sort stimuli into rough categories to facilitate fast, possibly life-saving, behavioral responses, and that contextual, individuating details could be added given more time. Bruner (1957) helped set the tone for this work by referring to early cognitive processing as "primitive categorization" and describing the overall perceptual process as "a 'bracketing' one, a gradual narrowing of the category placement of the object" (p. 130). In social psychology, this basic notion of contextualization was most readily translated into the domain of impression formation in an inter-racial context, where it has distinctly negative implications (e.g., Devine, 1989). Although the use of rigid categorical knowledge to quickly discriminate predator from prey might seem to be evolutionarily adaptive, in the modern world the use of such stereotypes would frequently lead to unwarranted discrimination (Dovidio, Kawakami, Smaok, Gaertner, 2008; cf. Arkes & Tetlock, 2004). On this view, evaluative and conceptual knowledge associated with a category in memory is automatically activated upon perception of a similar stimulus. It is the job of higher-level, controlled processing to monitor activated categorical knowledge and "correct" it, through the addition of situational information, as appropriate.

These early dual-process models of social-cognition were primarily focused on explicating how the bottom-up, monitoring-and-correction schema could describe impression formation (Brewer, 1988; Fiske & Neuberg, 1990), and racial prejudice in particular (Devine, 1989). For example, both Brewer (1988) and Fiske and Neuberg (1990) offered models in which the early impressions underlying automatic evaluations are based on categorical representations of a social target; controlled evaluations are then assumed to make use of individuating information available in the situation, thereby correcting erroneous stereotypical impressions, but only given time and effort. As Brewer (1988) put it:

> *Impressions are based on an active categorization process in which available 'person types' are matched to the information given about the new person. The search [for a matching category] is presumed to continue in an iterative … process, starting at the most general level of categorization and progressing to more specific subtypes, until an adequate fit is achieved. (p. 17).*

Similarly, Fiske and Neuberg (1990) assumed that "the sequential priority of processes [in impression formation] goes from category confirmation, to recategorization, to piecemeal integration of attributes" (p. 2). Devine (1989) applied this schema specifically to model the interaction between learned stereotypes and deliberate beliefs about social groups that was assumed to underlie prejudicial behavior:

> *Whereas high-prejudice persons are likely to have personal beliefs that overlap substantially with the cultural stereotype, low-prejudice persons have decided that the stereotype is an inappropriate basis for behavior or evaluation and experience a conflict between the automatically activated stereotype and their personal beliefs. The stereotype conflicts with their*

> *nonprejudiced, egalitarian values. The model assumes that the low-prejudice person must create a cognitive structure that represents his or her newer beliefs (e.g., belief in equality between the races, rejection of the stereotype, etc.). Because the stereotype has a longer history of activation (and thus greater frequency of activation) than the newly acquired personal beliefs, overt nonprejudiced responses require intentional inhibition of the automatically activated stereotype and activation of the newer personal belief structure. Such inhibition and initiation of new responses involves controlled processes. (p. 6)*

With the subsequent development of easily administered and fairly reliable indirect measurement procedures, a major goal of dual-process research became to explain the relation between responses on these procedures compared to responses on direct measurement procedures, which often appeared to be dissociated (e.g., Fazio & Olson, 2003; Greenwald & Banaji, 1995). Much of this research focused specifically on evaluative responses, leading to the development of a number of models describing the interaction between indirectly measured, automatic, *implicit* attitudes and directly measured, controlled, *explicit* attitudes in terms of the contextualization hypothesis (e.g., Cunningham, Zelazo, Packer, & Van Bavel, 2007; Fazio & Towles-Schwen, 1999; Hofmann et al., 2005; Rydell & McConnell, 2006; Wilson, Lindsey, & Schooler, 2000). Fazio's MODE model (Fazio, 2007; Fazio & Towles-Schwen, 1999), which had a strong influence on the field, is perhaps the prototypical contextualization model. The acronym MODE stands for "Motivation and Opportunity as DEterminants" of the influence of attitudes on behavior, where attitudes are specifically conceptualized as stable object-evaluation associations stored in memory. The automatic activation of these "summary" evaluations in memory is assumed to influence behavior unless the actor has both the motivation and opportunity to "correct" this influence by taking contextual details of the situation into account. As Fazio and Towles-Schwen (1999) described it, "The [automatic process] focuses upon preexisting attitudes and their accessibility from memory. This can be contrasted with a much more deliberative process in which the individual focuses not upon any preexisting attitude, but upon the raw data" (p. 99).

The contextualization hypothesis has received empirical support across several domains of social-cognitive research. For example, Dovidio and Gaertner (2004) developed their theory of aversive racism based on the idea that stable negative associations with outgroups in memory often conflict with egalitarian goals; once activated, these stable outgroup associations can be corrected only under controlled operating conditions (see also Dovidio, Kawakami, & Gaertner, 2002). Consistent with this idea, the more strongly outgroups are associated with negativity, the more cognitively depleting are cross-race interactions, presumably due to efforts to control the expression of these automatically activated associations (Richeson et al., 2003; Richeson & Shelton, 2003). The monitoring-and-correction processing schema has also been applied to self-regulation behaviors more generally, such that the influence of stable, automatically activated associations in memory can be "corrected" with respect to the current situational context, but only under conditions of control (e.g., Hofmann, Friese, & Strack, 2009). In addition, both self-esteem and the self-concept have been understood in these terms, with stable self-associations in memory influencing explicit self-descriptions unless controlled cognition is able to qualify these self-stereotypes (e.g., Greenwald & Farnham, 2000; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003; Nosek, Banaji, & Greenwald, 2002).

Although the contextualization hypothesis appears to be consistent with a wide range of empirical evidence, it also faces significant challenges. First, evidence for top-down influences, in which higher-level processing directly affects lower-level processing, challenges the bottom-up flow of processing implied by the hypothesis. For example, Peters and Gawronski (2011a, Experiment 2; see §2.2) found that being motivated to perceive oneself as extraverted (or introverted) produced congruent changes in explicit self-knowledge; significantly, these explicit changes mediated changes in the activation of implicit self-knowledge, presumably due to the top-down influence of a biased search for relevant memories. Similarly, Whitfield and Jordan (2009) found that reading behavioral descriptions of social targets (e.g., "Dan is rude to his mother") influenced explicit evaluations of the targets, and that these changes in explicit evaluations mediated changes in implicit evaluations of the targets. This mediation pattern was observed for both novel and familiar targets, suggesting that top-down processing can underlie both formation and change of evaluations in memory. Such findings are inconsistent with the contextualization hypothesis, which views higher-level processing essentially as a downstream integrator and modifier of automatically activated knowledge; from the perspective of this hypothesis, higher-level processing should be able to "correct" activated knowledge at the time of behavioral expression, but should not have a retroactive influence on knowledge stored in memory.

A second, more trenchant, challenge to the contextualization hypothesis comes from accumulating evidence that lower-level processing can be quite sensitive to situational context, often even more so than higher-level processing. In fact, empirical evidence for the context-insensitivity of implicit cognition is surprisingly rare (e.g., Foroni & Mayr, 2005; Gregg, Seibt, & Banaji, 2006; Peters & Gawronski, 2011b). In contrast, implicit cognition has been found to be sensitive to situationally active goals (Ferguson & Bargh, 2004; Foroni & Mayr, 2005), visual context (Gawronski, Rydell, Vervliet, & De Houwer, 2010), social roles and status (Barden, Maddux, Petty, & Brewer, 2004; Dasgupta & Greenwald, 2001; Sinclair, Lowery, Hardin, & Colangelo, 2005; Wittenbrink, Judd, & Park, 2001), directed imagination (Blair, Ma, & Lenton, 2001), and the physical environment (Cesario, Plaks, Hagiwara, Navarrete, & Higgins, 2010; Schaller, Park, & Mueller, 2003).

Strong evidence for the context-sensitivity of lower-level processing is difficult to reconcile with the contextualization hypothesis and the bottom-up, monitoring-and-correction models it informs. As a specification of the cognitive covariation thesis, the contextualization hypothesis seeks to carve cognition into two kinds of process that are distinguished principally by their sensitivity to situational context. Thus, lower-level processing is assumed to involve stable, categorical knowledge rooted in memory (i.e., stereotypes), whereas higher-level processing is assumed to involve flexible, individuated knowledge sensitive to situational demands. Faced with contradictory evidence, some proponents of the contextualization hypothesis have attempted to defend it through post-hoc adjustments to their models that would allow lower-level processing to be more context-sensitive. Fazio (2007), for example, has defended his MODE model by arguing that evidence for the context-sensitivity of lower-level processing can be explained by a change in how the object of evaluation is construed rather than by a change in the evaluation associated with the original object. Specifically, Fazio argues that because most objects are multiply categorizable, the evaluation of an object depends on how it is categorized *within a given context*; yet blurring the distinction between object and context weakens the conceptual foundation of the MODE model (i.e., the stability of object-evaluation associations in memory), making it difficult to distinguish from a more constructivist approach in which the "object of evaluation" is a unique construction based on the present context (e.g., Ferguson & Bargh, 2007; Schwarz, 2007). Ultimately, to the extent such post-hoc adjustments are made, the key distinction that makes these dual-process models empirically interesting (i.e., the contextualization hypothesis) is weakened, and the models become increasingly difficult to falsify.

1.2.2 The validation hypothesis

1.2.2.1 Basic assumptions

In distinguishing between different kinds of cognition, the contextualization hypothesis depends centrally upon the notion of compositionality, such that higher-level processing combines lower-level types, as the situation demands, to produce individuated tokens. From this perspective, higher-level processing serves a monitoring-and-correction function by contextualizing stable category knowledge to make it more situationally appropriate. The validation hypothesis builds upon the same monitoring-and-correction processing schema, but it does not draw the distinction between kinds of cognition in terms of representational complexity; thus, lower-level representations are assumed to be, in principle, as complex and situation-sensitive as higher-level knowledge. Instead, the distinction is drawn in terms of sensitivity to *truth-values*, such that the activation of lower-level knowledge and its influence on automatic behavior occurs independent of its perceived validity, but its influence on controlled behavior *does* depend on its perceived validity (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). The validation hypothesis therefore suggests that higher-level processing monitors lower-level processing by evaluating its validity and, if necessary, corrects it by rejecting activated knowledge perceived to be false.

According to Gawronski and Bodenhausen (2006), the perceived validity of activated knowledge is assumed to be a function of *situational consistency*—that is, consistency with the set of currently endorsed (i.e., previously validated) beliefs that make up an individual's current state of consciousness. Thus, as discussed below, perceived validity can be considered one particular aspect of the situational context. For example, whether or not the implication of an automatic negative reaction to a Black man (i.e., "I dislike black people") is perceived as valid will depend upon its consistency with other beliefs that are endorsed in the current situation (e.g., "Discrimination against minorities is wrong"; Gawronski, Peters, Brochu, & Strack, 2008).

A second point of difference between the validation and contextualization hypotheses concerns the direction of influence between lower-level and higher-level processing. The distinction between degrees of representational complexity assumed by the contextualization hypothesis has typically been understood as implying that the monitoring-and-correction schema operates in a bottom-up direction, whereby higher-level processing moderates the connection between lower-level processes and controlled behavior, but does not exert a retroactive influence on lower-level processes themselves. This is because the more complex, contextualized tokens represented at the higher level cannot, by assumption, be represented at the lower level. From the perspective of the validation hypothesis, however, top-down influences are more plausible since both processing levels are capable of equal representational complexity. Thus, direct manipulations of higher-level knowledge are expected to depend upon the activation of (indeed, to be identical with the activation of) corresponding lower-level knowledge. In contrast to the contextualization hypothesis, it is not the manipulation of higher-level knowledge *in general* that should have an asymmetric influence on higher-level and lower-level processes but specifically the manipulation of *perceived validity*. Thus, controlled recall of "true" knowledge (e.g., via logical inference) is assumed to rely upon the top-down activation of that knowledge at the lower level. On the other hand, the rejection of activated knowledge as "false" in higher-level processing should have no top-down influence on activated knowledge, since lower-level processing is assumed to be unaffected by perceived validity; for example, rejecting one's automatic negative reaction toward a Black man as invalid because it is inconsistent with one's egalitarian beliefs would not be expected to deactivate that negative association, which might then be expressed in automatic behaviors (Gawronski & Bodenhausen, 2006).

Finally, as with the contextualization hypothesis, the dual-process models based on the validation hypothesis also invoke the distinction between associative and rule-based operating principles in characterizing lower-level and higher-level processes (see Table 1.2). Thus, the automatic activation of knowledge is assumed to be a function of similarity between associations stored in memory, although the increased representational complexity of these associations makes associative processing more flexible. Likewise, higher-level processing is assumed to be a function of syllogistic rules, made possible by sensitivity to truth-values.[4]

The Validation Hypothesis	
Operating Conditions	
Automatic	Controlled
Unintentional	Intentional
Uncontrollable	Controllable
Resource-efficient	Resource-inefficient
Outside awareness	Within awareness
Operating Principles	
Implicit	Explicit
Insensitive to perceived validity*	Sensitive to perceived validity*
Associative	Rule-based

Table 1.2. The specification of the cognitive covariation thesis per the validation hypothesis.
*Denotes the key distinction between cognitive operating principles.

[4] In Gawronski and Bodenhausen’s (2006) APE model, higher-level processing is described as “propositional” rather than “rule-based” to emphasize its sensitivity to truth-values.

By invoking the associative/rule-based distinction, validation models carry normative connotations similar to those based upon the contextualization hypothesis, although these connotations are somewhat weaker given the increased context-sensitivity of lower-level processing. In fact, given these similarities and the common adherence to a monitoring-and-correction processing schema, it can be useful to view the validation hypothesis as a special case of the contextualization hypothesis in which sensitivity to one specific situational factor (i.e., perceived validity)—rather than to the situational context in general—distinguishes between kinds of cognitive process.

1.2.2.2 Empirical applications and challenges

The validation hypothesis as described above is realized most completely in Gawronski and Bodenhausen's (2006, in press) APE model and in Strack and Deutsch's (2004) Reflective-Impulsive Model (RIM). Both of these models view sensitivity to perceived validity as the key distinction between kinds of cognition. Moreover, viewed as special cases of contextualization models, the APE model and the RIM were specifically formulated to improve upon the empirical weaknesses of the monitoring-and-correction processing schema by allowing for lower-level context-sensitivity and top-down influences. Consequently, these models are consistent with the wealth of evidence for the context-sensitivity of implicit cognition and have led to a number of confirmed predictions regarding the top-down influence of explicit cognition on implicit cognition (e.g., Peters & Gawronski, 2011a; Whitfield & Jordan, 2009; see Gawronski & Bodenhausen, 2006, for a review). In addition, the APE model has been used to integrate various theoretical conceptions of prejudice, including Dovidio and Gaertner's (2004) theory of aversive racism, into a single conceptual framework based on the consistency between implicit evaluations and explicit beliefs (Gawronski et al., 2008).

Perhaps the most compelling evidence for the validation hypothesis, however, is the demonstration of an asymmetric influence of perceived validity on explicit and implicit cognition. For example, Gawronski and Strack (2004) employed the classic induced compliance paradigm from cognitive dissonance research to manipulate the availability of a situational explanation for writing a counter-attitudinal essay. When a situational explanation was available, writing the essay had no effect on participants' explicit or implicit evaluations of the topic. In contrast, when a situational explanation was unavailable, explicit evaluations of the topic became more positive, presumably to maintain consistency between attitudes and behavior. Critically, however, no effect of consistency was observed for implicit evaluations of the essay topic, which remained negative. Another line of evidence involves the asymmetric influence of negation on explicit and implicit cognition (Deutsch, Gawronski, & Strack, 2006; Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009). For example, Deutsch et al. (2006) found that the process of negating a valenced word (e.g., "no cockroach") requires a fixed amount of time that is unaffected by practice. Deutsch et al. interpreted this resistance to practice as evidence that negation is inherently a higher-level, rule-based process that cannot be automatized.

Despite its consistency with existing evidence and support for its novel predictions, the validation hypothesis has begun to face challenges as a specification of the cognitive covariation thesis. In particular, evidence has appeared that calls into question the distinction between automatic and controlled processes in terms of their sensitivity to truth-values. For example, Deutsch et al. (2009) found that evaluations can be quickly and efficiently negated when they are assessed with the AMP. As will be discussed in more detail below, the present research also suggests that negation exerts a powerful influence on the formation of implicit evaluations (Peters & Gawronski, 2011b; see §3). As with the contextualization hypothesis, proponents of the validation hypothesis have attempted to defend their models by weakening the specification of the cognitive covariation thesis. Thus, Gawronski and Bodenhausen (2009, in press) have responded to these empirical challenges by suggesting that automatic and controlled processes are not cleanly distinguished by their sensitivity to perceived validity (see §1.1.1.2). Once again, however, these efforts to defend the specification of the cognitive covariation thesis by weakening it—which amounts to denying the principle of covariation—make the models built upon the validation hypothesis more difficult to falsify. Instead, as argued above (§1.1.2), contradictory evidence may be more fruitfully used to make the specification of both covariation theses more empirically adequate. I will pursue this approach below by suggesting that it is not perceived validity *per se* that distinguishes between kinds of cognitive process, but rather *how* it is manipulated (see §4.3.1).

1.2.3 How many representations are there?

In pursuing this revised specification of the cognitive covariation thesis, the debate over the number of mental representations (i.e., memory traces) assumed to underlie dual-process cognition will be of central importance. This question was not addressed above because it is ultimately irrelevant for describing the contextualization and validation hypotheses: Existing single- and dual-representation versions of either hypothesis are theoretically interchangeable and empirically indistinguishable (Greenwald & Nosek, 2008). The reason for this is that existing dual-process models all assume that mental representation is compositional, such that complex thoughts are built from simple thoughts that possess a basic, discrete meaning and are themselves irreducible (Szabó, 2009). Contextualization models directly employ this assumption to distinguish between kinds of cognition, such that lower-level processes operate upon simple, decontextualized representations, which are combined by higher-level processes to create complex, contextualized representations. Existing validation models also assume that mental representation is compositional, but do not employ this assumption as a general distinction between kinds of cognition; instead, these models suggest it is the perceived validity, not the complexity, of mental representations that marks the distinction. In either case, then, existing dual-process models assume that mental representation is compositional, with the key consequence that representation at the higher level is assumed to be *reducible* to representation at the lower level. Thus, in dual-representation theories, which posit two distinct memory traces underlying implicit and explicit cognition, complex representations must be additive (non-emergent) combinations of simple representations; this means that a complex representation can in principle be completely reduced to simple representations, though the complex representation is assumed to exist as a distinct memory trace. In single-representation theories, on the other hand, complex representations are not assumed to exist as distinct memory traces, but just *are* combinations of simple representations.

Although it might seem natural to assume that distinct kinds of cognitive process operate upon distinct memory traces (especially when those processes are assumed to operate in parallel), the independent assumption that higher-level representations are reducible to lower-level representations leaves dual-representation models vulnerable to criticism on the grounds of parsimony. Specifically, dual-process models that posit two distinct memory traces, one of which is reducible to the other (e.g., Greenwald & Banaji, 1995; Rydell & McConnell, 2006; Wilson et al., 2000), introduce a redundancy that is not found in single-representation models (e.g., Fazio & Towles-Schwen, 1999; Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004).[5] The root problem is that the compositional relation between the two memory traces in dual-representation models undermines their classification as qualitatively different *kinds*: Because both higher-level and lower-level representations are built from the same components, they must in principle be reducible. Conversely, the parsimony critique implies that single-representation theories must assume that higher-level representation is completely reducible to lower-level representation, such that there is no loss of information when knowledge moves across levels; thus, these models are committed to a single kind of mental representation.

The parsimony critique is valid when targeting existing dual-representation models: Compositionality implies the reducibility of one kind of representation to another, creating a theoretical redundancy. As a critique of dual-representation theories *in general*, however, this argument begs the question by assuming that there is only one way an "object" can be mentally represented—that is, compositionally. This assumption reflects an entrenched linguocentrism in psychology, whereby natural language, which is

---

[5] Although dual-representation models in social-cognition generally avoid explicit statements of representational assumptions, these can be inferred from how implicit cognition is characterized. For example, Greenwald and Banaji (1995) define an implicit construct as an "introspectively unidentified (or inaccurately identified) trace of past experience" (p. 5), but do not claim that these constructs are unidentifiable *in principle*. Consistent with this interpretation, Greenwald and Banaji assume that the key empirical property of indirect measurement procedures is that they do not alert the participant to what is being measured, implying that it is possible for implicit constructs to be explicitly represented. Echoing Fazio's (1995) definition of an attitude as an object-evaluation association, Wilson et al. (2000) state that implicit and explicit attitudes are "summary evaluations that can be based on a variety of sources of information" (p. 107); thus, Wilson et al. claim that implicit and explicit attitudes do not differ in terms of how or what information is represented, but in the extent to which the attitude has become "ingrained" or automatized. Rydell and McConnell (2006) take a similar approach, assuming that all attitudes are object-evaluation associations, and distinguishing between explicit and implicit attitudes in terms of how they are learned. In each of these cases, the authors appear to assume that implicit representations can in principle be represented explicitly, implying their reducibility.

generally assumed to be compositional (Fodor, 2007; Szabó, 2009), is treated as the paradigm example of cognitive information processing. Consequently, the principles by which language operates are expected to inform our understanding of cognition:

> *Since the birth of cognitive science,* language *has provided the dominant theoretical model. Formal cognitive models have taken their structure from the syntax of formal languages, and their content from the semantics of natural language. The mind has been taken to be a machine for formal symbol manipulation, and the symbols manipulated have assumed essentially the same semantics as words of English. (Smolensky, 1988, p. 4, emphasis in orginal)*

This reasoning reached its apex in Fodor's (1975) language of thought hypothesis, which claimed that cognition can only be understood as computation—that is, as the rule-based manipulation of discrete symbols (see also Fodor & Pylyshyn, 1988). Opposition to the classical, linguocentric view of mental representation grew out of evidence for a distinction between verbal and imagistic cognitive processing (e.g., Kosslyn, 1980; Paivio, 1971) as well as advances in neural network modeling techniques that provided a plausible explanation for how non-computational cognition might be implemented. These *connectionist* models demonstrated that weighted networks of densely interconnected nodes could simulate many aspects of human cognition (Rogers & McClelland, 2004). Critically, however, these models do not require an assumption of compositionality, such that individual nodes would each possess a discrete meaning or represent a discrete object. Instead, they can also be constructed such that meaning is found only in overall patterns of activated nodes (Chrisley, 1998; Smolensky, 1988), analogous to the way in which the pattern of pixels on a monitor can be meaningful without any one pixel carrying a discrete meaning. This holistic, *distributed* style of representation is indeed a qualitatively different kind of mental representation compared to the classical, *localist* style of representation, and as such it provides a basis for dual-representation theories that are not vulnerable to the parsimony critique: Because distributed representations are non-compositional, they cannot be reduced to localist, compositional representations (and vice versa). Loosely speaking, this is why a bitmap image of a scene and a verbal description of the same scene will inevitably be non-equivalent—because they carry different kinds of information. Thus, as Smolensky (1988) noted, localist descriptions of distributed cognitive processing will inevitably be "incomplete (describing only certain aspects of the processing) or informal (describing complex behaviors in, say, qualitative terms) or imprecise (describing the performance up to certain approximations or idealizations…)" (p. 7).

In fact, dual-process models of social-cognition frequently appeal to the distinction between classical and connectionist architectures in specifying the cognitive covariation thesis (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004), but for very particular reasons. First, connectionist architectures provide a mechanism for similarity-based associative processing via the phenomenon of pattern matching, in which the activation of one pattern of nodes facilitates the activation of overlapping (and hence similar) patterns without any need for rule-based, symbolic computation (Smith, 1996). Second, connectionist architectures provide a plausible implementation-level account of associative processing, as they are inspired by the networks of neurons that make up the brain (Smolensky, 1988).

Most appeals to connectionist architectures in social-cognition, however, ignore the further possibility of non-compositional, distributed representation, and instead treat connectionist architectures simply as providing an associative (as opposed to rule-based) mechanism for processing compositional, localist representations. For single-representation models, adhering to a localist version of connectionism makes sense, because these models must assume that there is only one kind of mental representation. Indeed, single-representation theorists are admirably clear on this point. For example, Fazio (2007, p. 612) argues that the object-evaluation associations underlying his MODE model would be represented equivalently in either classical or connectionist architectures. Likewise, both Gawronski and Bodenhausen's (2006, in press) APE model and Strack and Deutsch's (2004) RIM appeal to connectionist networks underlying lower-level associative processing, but conceive of these networks in terms of discrete, localist nodes. For example, Strack and Deutsch (2004, Figure 3) describe the activation of nodes in a network representing the discrete concepts PERSON, ELDERLY, SLOW, etc., which can be directly combined into truth-evaluable propositions by higher-level processing.

The commitment to a single, localist kind of representation becomes less tenable, however, in light of empirical evidence suggesting that mental representation is not always discrete and categorical (Medin, 1989; Murphy, 2002; Rogers & McClelland, 2004). In summarizing this literature, Medin (1989) pointed to three critical findings: The inability of lay persons and experts alike to specify core feature sets for concepts; typicality effects, whereby some objects are considered “better” examples of a concept than others (e.g., an orange is a better exemplar of the concept FRUIT than a tomato); and unclear cases, whereby appeal to defining features seems to be unhelpful (e.g., does a rug count as instance of the concept FURNITURE?). The ubiquity of these findings suggests that category representations in general are not sets of necessary features, but are rather clusters of (sub-featural) properties that tend to covary—in other words, they are more like distributed patterns than discrete definitions. Based on such evidence, it seems that the question begged by the parsimony critique of dual-representation theories is not trivial. If different kinds of process are assumed to operate upon genuinely different kinds of representation (i.e., localist vs. distributed), the distinction between kinds of cognitive process will be, at least partly, a result of the distinction between kinds of representation; therefore, the theoretical properties distinguishing kinds of representation should lead to testable predictions about how to distinguish kinds of process, and in this way should inform the specification of the cognitive covariation thesis.

Below, I will explore the possibility of a dual-process theory that combines localist and distributed representations, and thereby takes full advantage of the insights from connectionist modeling (see Conrey & Smith, 2007, Smith & Conrey, 2007, for initial steps in this direction). As I will discuss, the possibility of non-compositional, distributed representation and its relation to compositional, localist representation may provide a solution to the challenges facing the contextualization and validation hypotheses.

1.2.4 Summary

In discussing the problem of specification in dual-process research, I have focused on how to specify the relation between the operating conditions and the operating principles of cognition—that is, the cognitive covariation thesis. Any such specification constitutes a hypothesis about how to carve cognition into distinct kinds of process. The two major hypotheses in social-cognitive psychology are both built upon a monitoring-and-correction processing schema in which knowledge automatically activated in lower-level processing is monitored relative to some criterion and corrected, given the conditions of control, by higher-level processing. In the case of the contextualization hypothesis, the criterion is situational appropriateness, such that higher-level processing functions to qualify automatically activated, stable category knowledge with respect to situational demands. The key principle distinguishing automatic and controlled processes is therefore assumed to be sensitivity to the situational context. In the case of the validation hypothesis, the criterion is situational consistency, such that higher-level processing functions to evaluate the validity of automatically activated knowledge with respect to currently endorsed beliefs. From this perspective, the key principle distinguishing automatic and controlled processes is assumed to be sensitivity to perceived validity.

Both hypotheses have received empirical support but also face significant challenges. The contextualization hypothesis does not readily predict top-down influences of controlled processing on automatic processing, and it is contradicted by evidence of the context-sensitivity of the latter. The validation hypothesis is consistent with both lines of evidence, but is itself challenged by findings suggesting that automatic processing can be sensitive to perceived validity. In response, I have suggested that the specification of the cognitive covariation thesis may need to account for distinct kinds of mental representation (i.e., localist vs. distributed) underlying kinds of cognitive process. This distinction has been largely ignored or misconstrued in existing dual-process models, which assume the compositionality of mental representation and thus that representations at one level are reducible to those at the other. If this is not the case, then understanding the relation between these (genuinely) different kinds of mental representation may be critical to the specification of the cognitive covariation thesis.

1.3 The unity problem

Most research and theorizing in the dual-process literature has focused on the specification problem and, to a lesser extent, the measurement problem. Yet there remains a more fundamental question that is rarely addressed: Why would cognition be divided into distinct kinds of cognition in the first place? And why would that division take the particular shape it does? Samuels (2009) refers to this as the unity problem, and he discusses it with respect to the common “cluster” approach to specifying the cognitive covariation thesis, in which clusters of operating principles correlate with different clusters of operating conditions (e.g., the received view of associative, similarity-based processing is that it is resource-efficient, unintentional, uncontrollable, and less accessible to conscious awareness). For Samuels, the unity problem follows on the heels of the specification problem: That is, once the cognitive covariation thesis has been specified, one can ask *why* this particular cluster of operating principles correlates with this particular cluster of operating conditions. For example, why is associative processing inherently less accessible to conscious awareness? Why is rule-based processing more controllable and less resource-efficient?

In fact, despite its popularity in the social-cognitive literature, the cluster approach to the specification problem involves dangerous theoretical excess, since there is no a priori reason that specification of the cognitive covariation thesis requires paired clusters of principles and conditions rather than a single key distinction to characterize each. Of course, the cluster approach to specification would seem to have the virtue of making the cognitive covariation thesis more falsifiable, because it makes stronger empirical claims. For example, on the received view, *all* of the functional properties of automaticity should be perfectly coincident, and should always correspond to the operating principles of implicit cognition. These are indeed strong empirical claims, and are readily disconfirmed by evidence that these clusters are not aligned (e.g., Gawronski et al., 2007; Moors & De Houwer, 2006; see Keren & Schul, 2009). The cluster approach to specifying the cognitive covariation thesis survives, however, because in practice these clusters are treated more like heuristics for deriving predictions rather than theoretical claims, making them moving targets that are difficult to falsify. Consequently, disconfirming evidence is more easily attributed to the failure of the psychometric covariation thesis—that is, due to the failure of indirect and direct measurement procedures to select purely automatic and controlled cognitive processes. As argued above, however (see §1.1.2.3), the process purity of measurement cannot be abandoned without compromising the ability to draw inferences about cognitive processes from observed behavior. For this reason, restricting specification of the cognitive covariation thesis to single distinctions between operating conditions and operating principles is not only theoretically simpler but empirically safer, as there is less room to maneuver around disconfirming evidence.

This has been the approach taken above, in which I sought to identify the *key* distinctions between operating conditions and operating principles that characterize the two major attempts to specify the cognitive covariation thesis. Thus, the contextualization hypothesis characterizes automatic processing as being comparatively insensitive to the situational context (Table 1.1), whereas the validation hypothesis characterizes automatic processing as being comparatively insensitive to perceived validity (Table 1.2). An advantage of these one-dimensional distinctions is that the unity problem becomes more tractable: Rather than explaining why clusters of operating principles covary with clusters of operating conditions in monolithic *systems* (cf. Keren & Schul, 2009), the challenge is to explain why a specific principle characterizes cognitive processing operating under a specific condition. Based on the earlier characterization of both major hypotheses in terms of a monitoring-and-correction processing schema, the question is therefore *why* early processing would be comparatively insensitive to the situational context (either in general or with respect to situational consistency in particular).

Although any response to the unity problem will almost certainly be speculative, it nevertheless imposes an important constraint on attempts to specify the cognitive covariation thesis—that is, that the specification respect known neurological, and ultimately phylogenetic, distinctions between cognitive processing capacities (Evans, 2008). At minimum, then, any specification of the cognitive covariation thesis should be supported by a prima facie valid story about why cognitive processes are distinguished into the particular kinds specified. Conversely, dual-process theories unable to supply a reasonable explanation for the distinction in cognitive processes they propose should lose a measure of support. Below, I will briefly consider three such explanations of the monitoring-and-correction processing schema, which forms the basis for both the contextualization and validation hypotheses, and argue that all three are inadequate.

1.3.1 The architectural explanation

To the extent that they provide an explicit response to the unity problem, dual-process theories of social-cognition have focused on implementation-level analyses. As discussed earlier, many dual-process models have appealed to the distinction between classical and connectionist cognitive architectures to ground the distinction between controlled and automatic processes (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004). The general idea is that classical architectures provide the capacity to flexibly qualify and "correct" stable knowledge automatically activated in connectionist networks, either with respect to the situational context in general or situational consistency in particular. Such uses of the classical/connectionist distinction, however, ignore the deeper implications of localist vs. distributed representation—specifically, that distributed representations are inherently no less flexible and context-sensitive than localist representations (and potentially even more so; Smolensky, 1988). Nor are connectionist networks in principle incapable of encoding truth-values as a property of learned information, though distributed representations of truth-values may be more probabilistic than binary (cf. Osherson & Smith, 1981).

1.3.2 The neurophysiological explanation

A slightly different approach is to ground the distinction within two physiologically distinct memory systems in the brain (e.g., Lieberman, 2007), usually divided into subcortical and cortical systems. The cortical system is assumed to have exclusive access to serial working memory, which accounts for both its capacity for deliberate, syllogistic reasoning and its relative slowness and inefficiency. As critics have pointed out, however, dependence on working memory may provide an explanation for the properties of controlled processing but sheds little light on the nature of automatic processing (Evans, 2008; Samuels, 2009). In response, the neurophysiological distinction is sometimes linked to the architectural distinction, such that the subcortical system is assumed to implement connectionist processing while the cortical system implements classical (rule-based) processing (Smith & DeCoster, 2000). As just discussed, however, the classical/connectionist distinction provides no inherent justification for the distinction between early and late processing in terms of sensitivity to the situational context when both architectures are assumed to operate upon localist representations.

1.3.3 The phylogenetic explanation

Nevertheless, it is often assumed that the neurophysiological distinction reflects a deeper phylogenetic distinction, such that the subcortical system is evolutionarily older than the cortical system (Evans, 2008; Stanovich, 2004). Ultimately, the appeal to phylogeny boils down to an evolutionary story that has (often tacitly) provided the inspiration for the monitoring-and-correction processing schema (see §1.2.1.1). The assumption is that, to be evolutionarily adaptive, the early cognition underlying fast responses must quickly match stimuli in the environment with rough, stable categories (e.g., predator vs. prey). The incorporation of contextual information, either to individuate the early stimulus representation or to assess the situational consistency of the early response, is thus a luxury afforded by time. Moreover, it is assumed that only higher animals have developed the capacity for this downstream, controlled processing (which is an important source of the normative connotations that continue to color dual-process theorizing in the monitoring-and-correction tradition).

The problem with this evolutionary story is that it doesn't make much sense: To be evolutionarily adaptive, early responding needs to be flexible and context-sensitive, not rigid and context-insensitive (Schwarz, 2007). In the complexity of the ecological environment, the discrimination of a stimulus as predator vs. prey would frequently fail if the situational context were not taken into account; rather, early cognition needs to do its best with incomplete and varying contextual cues to determine what a stimulus is. This supposition is strongly supported by the evidence reviewed in §1.2.1.2, which demonstrates that that early affective reactions to a stimulus are moderated by the context in which it is perceived (e.g., Cesario et al., 2010; Schaller et al., 2003).

1.3.4 Summary

In the end, the monitoring-and-correction processing schema appears to lack a reasonable explanation for the distinction between cognitive processes in terms of sensitivity to the situational context. Architectural, neurophysiological, and phylogenetic explanations have all been proposed, but none stands up to analysis (see also Samuels, 2009). Although this is not a devastating flaw, it does cast doubt on the accuracy of the monitoring-and-correction processing schema and, by implication, the contextualization and validation hypotheses built upon it.

1.4 Overview of present research

Above, I discussed three problems that must be addressed in any attempt to carve cognition into distinct kinds of process. Based on this discussion, the received view of how to carve cognition (Figure 1.2) appears to face serious difficulties. To begin with, the measurement problem cannot be solved by rejecting the covariation theses in principle, as these are required for cognitivist research; instead, the lack of process purity indicates a need to draw cleaner distinctions between observation conditions, operating conditions, and operating principles. The two major attempts to specify the cognitive covariation thesis in social-cognitive psychology both face empirical challenges. Furthermore, these hypotheses lack reasonable explanations for why higher-level processing would be distinguished by greater sensitivity to the situational context (either generally or with respect to situational consistency in particular). These difficulties suggest that the underlying monitoring-and-correction processing schema, in which higher-level processes function to control or make sense of lower-level impulses, may provide a misleading metaphor for carving cognition.

The present research has implications for each of these problems. In both sets of studies, the observed relations between implicit and explicit cognition cast doubt on current assumptions about how these processes interact. The first manuscript investigates the dynamics underlying self-concept change in terms of the interaction between implicit and explicit self-knowledge. This research demonstrates that implicit and explicit cognition exert mutual bottom-up and top-down influences on each other, which is consistent with the validation hypothesis but inconsistent with the assumption of the contextualization hypothesis that higher-level processing functions merely as an integrator of automatically activated lower-level knowledge. The second manuscript investigates the mechanisms underlying evaluative learning, demonstrating that perceived validity can, under certain conditions, have an impact on the learning of both explicit and implicit evaluations. To the extent that implicit and explicit evaluations are assumed to be learned via different mechanisms, this research has two implications immediately relevant to the current discussion. First, it suggests a strong top-down effect of rule-based learning, which again contradicts the assumption of the contextualization hypothesis that the direction of processing is bottom-up. Second, it suggests that implicit cognition may be sensitive to perceived validity under certain conditions, which contradicts the assumption of the validation hypothesis that implicit cognition should be insensitive to perceived validity.[6]

In the final chapter, I will discuss the implications of this research for the questions raised above. The major goal will be to provide a more empirically accurate specification of the psychometric and cognitive covariation theses, which can help guide future research on dual-process cognition. Ultimately, I will argue that existing social-cognitive theories have identified important pieces of the dual-process puzzle, but have put them together in the wrong way. The present research points toward a different and potentially more accurate approach to fitting these pieces together.

[6] To prevent confusion, it should be noted that the research reported in §2 on self-concept change is framed in terms of the validation hypothesis, and the APE model in particular. Although the findings from these studies are consistent with this perspective, they contradict the assumptions of the contextualization hypothesis. Furthermore, these findings are also consistent with an alternative interpretation of the relation between higher-level and lower-level cognition developed later (see §4).

1.5 References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Erlbaum.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test?”. *Psychological Inquiry, 15*, 257-278.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5-22.

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

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828-841.

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

Bosson, J., K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind man and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631-643.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*, 123-152.

Cesario, J., Plaks, J. E., Hagiwara, N., Nvarrete, C. D., & Higgins, E. T. (2010). The ecology of automaticity: How situational contingencies shape action semantics and social behavior. *Psychological Science*, *21*, 1311-1317.

Chomsky, N. (1959). Review of *Verbal Behavior*. *Language*, *35*, 26-58.

Chrisley, R. (1998). Non-compositional representation in connectionist networks. In Niklasson, L., Boden, M., & Ziemke, T. (Eds.), *ICANN 98: Proceedings of the 8th International Conference on Artificial Neural Networks*.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469-487.

Conrey, F. R., & Smith, E. R. (2007). Attitude representation: Attitudes as patterns in a distributed, connectionist representational system. *Social Cognition*, *25*, 718-735.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163-170.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*, 736-760.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 811-814.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. Wiers & A. Stacy (Eds.), *Handbook of implicit cognition and addiction* (1st ed., pp. 11-28). Thousand Oaks, CA: Sage.

De Houwer, J. (2008). Comparing measures of attitudes at the functional level and procedural level: Analysis and implications. In R. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures*. New York: Psychology Press.

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

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

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385-405.

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

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

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 1-52). San Diego: Elsevier.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62-68.

Dovidio, J. F., Kawakami, K., Smoak, N., & Gaertner, S. L. (2008). The roles of implicit and explicit processes in contemporary prejudice. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 165-192). New York: Psychology Press.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247-282). Hillsdale, NJ: Erlbaum.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 736-760.

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297-327.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97-116). New York: Guilford.

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

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 216-246). New York: Guilford.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York: Academic Press.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105-116). Malden, MA: Blackwell.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single reading of a short scenario. *Psychonomic Bulletin & Review*, *12*, 139-144.

Freud, S. (1927). *The ego and the id*. London: Hogarth Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731.

Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences, 32,* 207-208.

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

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

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24,* 218-225.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2*, 181-193.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34,* 648-665.

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

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

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist*, *47*, 766-779.

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

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022-1038.

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

Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65-82) . Hillsdale, NJ: Erlbaum.

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

Hofmann, W., Friese, M., & Strack, F. (2009) Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, *4*, 162-176.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369-1385.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, *16*, 335-390.

Hofmann, W., & Wilson, T. D. (2010). Consciousness, introspection, and the adaptive unconscious. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

Jacoby, L. L. (1991). A process-dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Henry Holt & Co.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, *85*, 969-978.

Keren, G., & Schul, Y. (2009) Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533-550.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge: Harvard University Press.

Kruglanski, A. W., Erb, H., Pierro, A., Mannetti, L., & Chun, W. Y. (2006). On parametric continuities in the world of binary either ors. *Psychological Inquiry*, *17*, 153-165.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: Impact of unreliability on the replicability of experimental findings involving implicit measures. *Personality and Social Psychology Bulletin*, *37*, 570-583.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259-289.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*, 1469-1481.

Murphy, G. L. (2002). Typicality and the classical view of categories. In G. L. Murphy, *The big book of concepts* (pp. 11-40). Cambridge, MA: MIT Press.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*, 44-59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265-292). New York: Psychology Press.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ratliff (Ranganath), K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36-88.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35-58.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181-192.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*, 272-314.

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

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

Peters, K. R., & Gawronski, B. (2011a). Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology, 47,* 436-442.

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

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*, 386-396.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, *6*, 1323-1328.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287-290.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: The MIT Press.

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

Samuels, R. (2009). The magical number two, plus or minus: Dual process theory as a theory of cognitive kinds. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Schaller, M., Park, J. J., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin*, *29*, 637-649.

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

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry*, *17*, 173-184.

Sherman, J. W. (2008). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 391-426). New York: Psychology Press.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology*, *89*, 583-592.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Smith, E. R. (1996). What do connectionism and social psychology have to offer each other? *Journal of Personality and Social Psychology*, *70*, 893-912.

Smith, E. R., & Conrey, F. R. (2007). Mental representations are states, not things: Implications for implicit and explicit measurement. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 247-264). New York: Guilford.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108-131.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1-74.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: The University of Chicago Press.

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

Szabó, Z. G. (2009). Compositionality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2009 ed.). Retrieved from http://plato.stanford.edu/archives/spr2009/entries/compositionality/

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology, 45,* 748-759.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101-126.

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

2 Mutual Influences between the Implicit and Explicit Self-Concepts: The Role of Memory Activation and Motivated Reasoning[7]

Historically, the self-concept has been understood as the collection of things we believe about ourselves. The use of recently developed techniques for indirectly measuring mental contents, however, has suggested that an individual's self-concept can differ depending on how it is measured—specifically, depending on whether information about the self is *explicit* or *implicit* in behavioral responses on the measurement procedure (cf. De Houwer, 2006). Thus, measures of the "explicit" self-concept, typically assessed via self-report, have been shown to diverge from measures of the "implicit" self-concept, typically assessed via performance-based measures (e.g., Greenwald & Farnham, 2000; see Schnabel & Asendorpf, 2010, for a review). Measures of the implicit self-concept predict behavior above and beyond measures of the explicit self-concept (Asendorpf, Banse, & Mücke, 2002; Back, Schmukle, & Egloff, 2009), and this incremental validity appears to derive from a difference between the conditions under which the two types of self-information influence behavior. For example, Asendorpf et al. (2002) found evidence for a double dissociation between explicit and implicit self-concepts, such that the explicit self-concept uniquely predicted controlled behaviors and the implicit self-concept uniquely predicted spontaneous behaviors. Moreover, it has been shown that discrepancies between the explicit and implicit self-concepts on a particular dimension (e.g., shyness) uniquely predict behaviors intended to reduce these discrepancies (Briñol, Petty, & Wheeler, 2006).

Based on these findings, it appears that measures of explicit and implicit self-concepts do tap different types of information about the self, and that these different self-conceptions can become dissociated. It is currently less clear, however, how these two self-conceptions may correspond. Based on theoretical perspectives that conceive of the self

---

[7] A version of this chapter has been published elsewhere (Peters & Gawronski) and has been used with permission of Elsevier Limited.

as an integrated system for facilitating adaptive behavior (Cross & Markus, 1990; Steele, 1988), it makes sense to expect these conceptions to be related. In fact, measures of the explicit and implicit self-concepts are typically correlated, suggesting a significant degree of correspondence. For example, Asendorpf et al. (2002), Briñol et al. (2006), and Back et al. (2009) observed correlations in the range of .30 to .40 between measures of the explicit and implicit self-concepts. A meta-analysis of correlations between self-report measures and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) also found that these assessments of the explicit and implicit self-concepts correlated at .21, more strongly than assessments of self-esteem and roughly equal to the overall correlation between self-report measures and the IAT across research domains (Hofmann, Gawronski, Gschwender, Le, & Schmitt, 2005). Thus, although explicit and implicit self-conceptions appear to be distinct, it is equally clear that they can correspond, raising the possibility that these two aspects of self-representation are related through processes of mutual influence.

The aim of the present research is to provide a framework for understanding these mutual influences by viewing the explicit and implicit self-concepts as distinct but interacting aspects of an individual's self-representation. Within this framework, measures of the implicit self-concept are assumed to reflect the momentary activation of specific self-associations in memory. Measures of the explicit self-concept are assumed to reflect validated self-beliefs, which are descriptive propositions about the self based on activated self-associations that are regarded as *true* by the individual. From this perspective, the explicit self-concept can be considered a "working" self-concept, in that it constitutes a continuously maintained network of beliefs about the self (Markus & Wurf, 1987). The implicit self-concept provides an online, context-sensitive source of activated information that substantiates, and potentially informs the revision of, this network of self-beliefs. The construction and maintenance of the working self-concept—a process we refer to as *self-construal*—is thus understood as a fundamentally epistemic enterprise involving the generation and validation of self-beliefs (Kruglanski, 1989; Quine & Ullian, 1970).[8]

This framework for relating the explicit and implicit self-concepts suggests two specific routes of influence underlying the general process of self-construal. First, a bottom-up, "data-driven" process of self-construal can occur when self-associations in memory are activated without the intention to revise the explicit, working self-concept. The increased accessibility of newly activated self-knowledge will then promote its incorporation into the explicit, working self-concept. Second, a top-down, "hypothesis-driven" process of self-construal can occur when the explicit, working self-concept is intentionally revised, which involves asserting the validity of a new propositional belief about the self (e.g., "I am extraverted"). To test this hypothesis about the self, autobiographical memory can be searched for relevant evidence; however, it is expected that this search will be biased toward activating confirmatory information that substantiates the asserted self-belief (Klayman & Ha, 1987; Kunda, 1990). In either of these cases, the process of self-construal should produce correspondence between the explicit and implicit self-concepts. The key difference is that during bottom-up self-construal, change in the implicit self-concept is expected to mediate change in the explicit self-concept, whereas during top-down self-construal, the reverse mediation is expected.

The following two experiments were designed to test these predictions concerning the mutual influences between the explicit and implicit self-concepts. In Experiment 1, self-associations in memory were activated independently of the intention to revise the explicit, working self-concept as a test of bottom-up self-construal. In Experiment 2, participants were motivated to revise their working self-concepts directly as a test of top-down self-construal. By relating the explicit and implicit self-concepts together within an

---

[8] In line with the broader use of the term *construal* in the social-cognitive literature, we use the term *self-construal* to refer to a general process of constructing beliefs based on momentarily accessible information. Previous use of the term *self-construal* to refer to the influence of culture on self-definition (e.g., Markus & Kitayama, 1991) can thus be understood within the present framework as self-construal in a particular content domain, whereby cultural factors influence the accessibility and desirability of specific self-information (e.g., independent versus interdependent self-characteristics).

overall framework of self-construal, this research promises to clarify the connection between these two aspects of self-representation.

2.1 Experiment 1

The first experiment tested the proposed account of bottom-up self-construal by asking participants to recall autobiographical memories relating to a specific personality trait (ostensibly as part of a study investigating the relation between personality styles and autobiographical memory). The recall task was intended to activate specific self-associations in memory independent of the intention to revise the working self-concept, thus initiating a process of bottom-up self-construal. To ensure that most participants would possess relevant memories and that the revision of the working self-concept implied by activated self-associations would not be resisted, it was necessary to manipulate a relatively broad and malleable domain of self-knowledge. Toward that end, the trait dimension of extraversion-introversion was identified as sufficiently fluid to ensure that most participants would be willing and able to see themselves as more or less extraverted (Sanitioso, Kunda, & Fong, 1990).

The current experiment tested three specific predictions derived from the framework of self-construal outlined above. First, the measure of the implicit self-concept (in this case, a self-concept IAT; Greenwald & Farnham, 2000) was expected to reflect the selective activation of self-associations congruent with the personality trait (i.e., extraversion or introversion) targeted in the memory recall task. Second, the measure of the explicit, working self-concept was expected to reflect the revision of beliefs about the self in line with the recalled memories, such that participants recalling extraverted (or introverted) memories would report more (or fewer) extraversion-related self-beliefs. Finally, it was predicted that these changes in the explicit, working self-concept would be mediated by changes in the activation of self-associations in the implicit self-concept, consistent with the proposed account of bottom-up self-construal.

2.1.1 Method

2.1.1.1 Sample and design

A total of 118 undergraduate students (80 women and 38 men) participated in a study on personality and autobiographical memory for course credit. The experimental design consisted of a single between-subjects factor with two conditions (Recalled Trait: Extraversion vs. Introversion). Order of the two dependent measures was counterbalanced across participants.

2.1.1.2 Memory recall task

Upon entering the lab, participants were seated at individual computer carrels and given informed consent documents to sign. Participants then began the memory recall task, which guided them through the process of sequentially recalling and describing two memories of their past behavior that they considered to be extraverted or introverted, according to the experimental condition. To encourage recalled behaviors to be interpreted as arising from the self rather than situational influences, participants were instructed to recall each memory using an observer's (as opposed to an actor's) visual perspective (Libby, Eibach, & Gilovich, 2005). For each memory, once the participant indicated that the requested memory had been recalled, a series of brief questions was asked to increase its vividness (e.g., "Can you see what your facial expression was?"; Libby et al., 2005). Participants were then asked to describe the recalled memory briefly in writing, again using an observer's visual perspective.

2.1.1.3 Measurement of implicit self-concept

A "self/extravert" IAT was used to assess the selective activation of trait-related self-knowledge following the memory recall task (see Appendix A for stimuli). The IAT compares reaction times to responses that pair a target (e.g., *me*) with an attribute (e.g., *extraverted*) against responses that pair the same target (e.g., *me*) with a complementary attribute (e.g., *introverted*). The resulting difference score provides a sample-relative index of the degree to which target-attribute associations are activated in memory. In the first block of the IAT, "me" and "not me" words had to be assigned to the categories *Me* (right) and *Not Me* (left). In the second block, extraversion and introversion words had to be assigned to the categories *Extravert* (right) and *Introvert* (left). In the third block, target and attribute trials were presented in alternating order, with “me” and extraversion words on the right and “not me” and introversion words on the left. In the fourth block, participants practiced categorizing only extraversion and introversion words with key assignments reversed. In the fifth block, target and attribute trials were again combined, with “me” and introversion words on the right and “not me” and extraversion words on the left. Blocks 1, 2, and 4 consisted of 20 trials, and blocks 3 and 5 consisted of 80 trials. The inter-trial interval was 250 ms. Following incorrect responses the word “ERROR!” was presented in the center of the screen for 1000 ms.

2.1.1.4 Measurement of explicit self-concept

A self-report rating scale was used to measure participants’ perceptions of their own personality traits, which are assumed to reflect the self-beliefs constituting the explicit, working self-concept. To mitigate demand effects following the memory recall task, participants were told that the researchers were “also interested in how a variety of personality dimensions influence recalled memories” and the scale was therefore presented as a general personality assessment. The scale consisted of six items relating to extraversion and six items relating to introversion (identical to the stimuli used in the IAT), along with six positively valenced filler items and six negatively valenced filler items (see Appendix B). The items were presented in an a priori randomized order and were rated on a 7-point scale.

All participants were debriefed at the completion of the experiment. None indicated suspicion of a link between the memory recall task and either of the dependent measures.

2.1.2 Results

2.1.2.1 Data preparation

An index of extraversion-related (vs. introversion-related) self-associations was calculated from responses in the IAT following Greenwald, Nosek, and Banaji’s (2003) *D*-600 algorithm (Cronbach’s $\alpha = .84$). Scores were calculated such that higher values reflect stronger associations between the self and extraversion (compared to introversion) in the implicit self-concept. An index of extraversion-related (vs. introversion-related) self-beliefs was calculated from the self-report scale by reverse-coding the six introversion-related items and computing the combined mean of the six extraversion-related items with the six reverse-coded introversion items (Cronbach's $\alpha = .92$). Higher scores therefore reflect a more extraverted (compared to introverted) explicit, working self-concept. The index of activated self-knowledge and the index of self-beliefs were significantly correlated, $r = .53$, $p < .001$.

2.1.2.2 Effects of memory recall task

Inspection of participants' written descriptions of recalled memories suggested that they complied with instructions to recall the requested extraversion- or introversion-related memories. Means and standard deviations for the two self-concept measures are presented in Table 2.1. As predicted, participants who recalled extraversion-related memories revealed significantly stronger associations between the self and extraversion (relative to introversion) on the IAT compared to participants who recalled introversion-related memories, $t(116) = 2.54$, $p = .01$, $d = 0.47$. Similarly, and again in line with predictions, participants who recalled extraversion-related memories reported significantly more extraverted (relative to introverted) self-beliefs than participants who recalled introversion-related memories, $t(116) = 1.97$, $p = .05$, $d = 0.37$.

	Experiment 1				Experiment 2			
	Recall Extraversion		Recall Introversion		Pro-Extraversion		Pro-Introversion	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Explicit Self-Concept	4.64	1.10	4.24	1.09	4.98	1.04	4.53	1.04
Implicit Self-Concept	0.61	0.56	0.35	0.53	0.61	0.45	0.44	0.52

Table 2.1. Means and standard deviations by condition for measures of explicit and implicit self-concepts in Experiments 2.1 and 2.2.

2.1.2.3 Mediation analysis

The third prediction tested in the current experiment was that the activation of trait-related self-associations in memory would mediate the effect of the recall task on the explicit, working self-concept, in line with the proposed account of bottom-up self-construal (see Figure 2.1). To test this prediction, self-report scores were simultaneously regressed on both the memory recall task and IAT scores (Baron & Kenny, 1986). The relation between IAT scores and self-report scores remained significant, $\beta = .51$, $t(115) = 6.31$, $p < .001$, but the effect of the recall task became non-significant, $\beta = .06$, $t(115) = 0.77$, $p = .45$. Thus, changes in the implicit self-concept fully accounted for changes in the explicit self-concept. The indirect effect of the memory recall task on self-report scores when IAT scores were included as a mediator was significant, Sobel's $Z = 2.38$, $p = .02$.

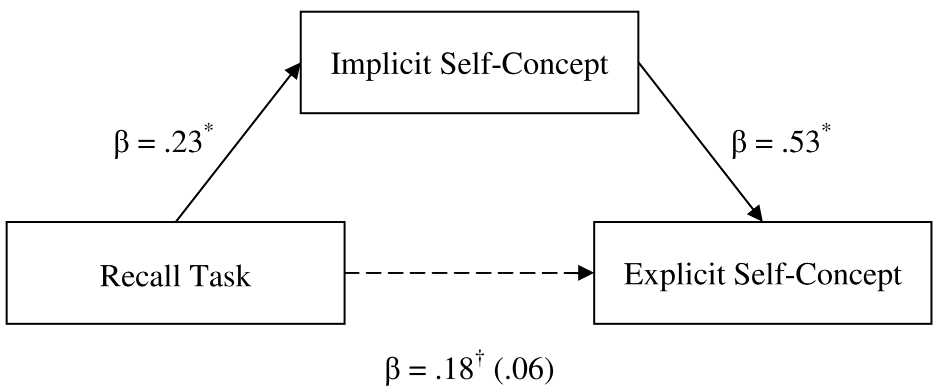

Figure 2.1. Mediation model tested in Experiment 2.1 (on the basis of Baron & Kenny, 1986). The indirect effect of the recall task on the explicit self-concept (mediated by the implicit self-concept) is statistically significant, Sobel's $Z = 2.38$, $p = .02$.
$^{\dagger}p = .05$; $^{*}p < .05$.

To rule out the alternative possibility of a top-down influence, we also tested the reverse mediation model, in which IAT scores were simultaneously regressed on both the memory recall task and self-report scores. In this mediation model, changes in self-report scores failed to account for the obtained effect on IAT scores, in that the memory recall task still had a marginally significant effect on IAT scores after controlling for self-report scores, $\beta = .14$, $t(115) = 1.76$, $p = .08$. Thus, changes in the implicit self-concept fully accounted for changes in the explicit self-concept, but not the other way around.

2.1.3 Discussion

The results of Experiment 1 provided support for all three predictions concerning bottom-up self-construal. The memory recall task selectively activated trait-related self-associations in the implicit self-concept and also led to congruent revision of the explicit, working self-concept. The mediation analysis supported the prediction that the effect of the recall task on the explicit self-concept would be mediated by activation of self-associations in the implicit self-concept. Consistent with this prediction, changes in self-associations fully accounted for changes in explicit self-beliefs, but not the other way around. Taken together, these findings suggest that differences in participants' explicit, working self-concepts between the memory recall conditions were due to the bottom-up integration of the self-beliefs implied by activated self-associations in memory.

2.2 Experiment 2

The second experiment was designed to test the proposed account of top-down self-construal. In contrast to bottom-up self-construal, in which the explicit, working self-concept is influenced independent of the intention to revise it, top-down self-construal begins with the intentional revision of the working self-concept. Thus, participants in the current experiment were asked to generate explanations for (fabricated) scientific findings linking either extraversion or introversion to successful life outcomes. By making a specific personality trait desirable, this manipulation was intended to motivate participants to assert the validity of the corresponding propositional belief about themselves (i.e., "I am extraverted" or "I am introverted"). Having marked this propositional belief as valid, participants were expected to treat it as a hypothesis to be tested by searching autobiographical memory for relevant evidence. Due to the confirmatory nature of hypothesis testing (Klayman & Ha, 1987), however, participants were expected to selectively activate self-associations in memory that would substantiate the asserted propositional belief (Sanitioso et al., 1990), leading to corresponding effects on the explicit and implicit self-concepts.

The current experiment tested three specific predictions derived from the proposed framework of self-construal. First, the measure of the explicit, working self-concept was expected to reflect the revision of beliefs about the self in line with the motivation manipulation, such that participants motivated to see themselves as extraverted (or introverted) would report more (or fewer) extraversion-related self-beliefs. Second, the measure of the implicit self-concept (a self-concept IAT) was expected to reflect the selective activation of self-associations congruent with the personality trait (i.e., extraversion or introversion) that participants were motivated to believe they possessed. Finally, it was predicted that these changes in the implicit self-concept would be mediated by changes in the explicit, working self-concept, consistent with the proposed account of top-down self-construal.

2.2.1 Method

2.2.1.1 Sample and design

A total of 148 undergraduate students (111 women and 37 men) participated in a study on personality and explanation styles for course credit. The experimental design consisted of a single between-subjects factor with two conditions (Desired Trait: Pro-Extraversion vs. Pro-Introversion). Order of the two dependent measures was counterbalanced across participants.

2.2.1.2 Motivation induction task

Upon entering the lab, participants were seated at individual computer carrels and given informed consent documents to sign. Participants then began the motivation induction task (adapted from Sanitioso et al., 1990). The task was framed as an investigation of how people generate explanations for scientific findings. As with Experiment 1, the current experiment manipulated self-perceptions along the extraversion-introversion dimension of personality. Participants were presented with a fabricated newspaper clipping reporting the findings of a recent scientific study comparing the benefits of extraverted personality traits with introverted personality traits (Appendix C). The clipping briefly described a study that found that extraversion leads to more academic and job success than introversion (or vice versa in the Pro-Introvert condition). After reading the clipping, participants were asked to write down two brief explanations for the observed relationship between extraversion (or introversion) and positive life outcomes. Generating the two explanations was intended to strengthen the manipulation and reinforce the cover story. Because most people presumably desire to see themselves as successful in life, the newspaper clipping was expected to motivate participants to hypothesize that they themselves possessed the personality trait that contributed to positive life outcomes, thereby initiating a process of top-down self-construal.

2.2.1.3 Dependent measures

The measures of the implicit and explicit self-concepts were identical to those used in Experiment 1.

2.2.1.4 Control measure and manipulation check

A manipulation check was included to ensure that the motivation induction task influenced the desirability of extraverted and introverted personality traits. For three of the extraversion-related and three of the introversion-related trait words used in the measurement of the explicit self-concept, participants were asked to indicate on a 7-point scale how much that trait contributed to success after university.

All participants were debriefed upon completion of the experiment. None indicated suspicion of a link between the motivation induction task and either of the dependent measures.

2.2.2 Results

2.2.2.1 Data preparation

Indices of extraversion-related (vs. introversion-related) self-associations (Cronbach's α = .76) and of extraversion-related (vs. introversion-related) self-beliefs (Cronbach's α = .91) were calculated as described in Experiment 1. The two indices were significantly correlated, $r = .46$, $p < .001$. For the manipulation check, an index of the degree to which extraversion vs. introversion contributes to positive life outcomes was calculated by reverse-coding the three introversion-related items and computing the combined mean with the three extraversion-related items (Cronbach's $\alpha = .85$). Higher scores thus reflect increased desirability of extraversion compared to introversion.

2.2.2.2 Manipulation check

The motivation induction task led participants in the Pro-Extravert condition to report that extraversion was more desirable ($M = 5.82$, $SD = 0.75$) than participants in the Pro-Introvert condition ($M = 5.10$, $SD = 0.78$), $t(146) = 5.75$, $p < .001$, $d = 0.95$. Thus, given that most people desire positive life outcomes for themselves, it is reasonable to assume that the induction task indeed motivated participants to perceive themselves as possessing more extraverted or introverted qualities, according to the experimental condition.

2.2.2.3 Effects of motivation induction task

Means and standard deviations for the primary measures are presented in Table 2.1. Participants in the Pro-Extravert condition revealed significantly stronger associations between the self and extraversion (relative to introversion) on the IAT compared to participants in the Pro-Introvert condition, $t(146) = 2.10$, $p = .04$, $d = 0.35$. Similarly, and again in line with predictions, participants in the Pro-Extravert condition reported significantly more extraverted (relative to introverted) self-beliefs than participants in the Pro-Introvert condition, $t(146) = 2.65$, $p = .01$, $d = 0.44$.

2.2.2.4 Mediation analysis

The third prediction tested in the current experiment was that asserting the validity of a propositional belief within the explicit, working self-concept would initiate a biased search through memory to activate substantiating self-associations, thereby mediating the effect of the motivation induction on the implicit self-concept (see Figure 2.2). To test this prediction, IAT scores were simultaneously regressed on both the motivation induction and self-report scores (Baron & Kenny, 1986). The relation between self-report scores and IAT scores remained significant, $\beta = .45$, $t(145) = 5.97$, $p < .001$, but the effect of the motivation induction became non-significant, β = .08, $t(145) = 1.00$, $p = .32$. Thus, changes in the explicit self-concept fully accounted for changes in the implicit self-concept. The indirect effect of the motivation induction on IAT scores when self-report scores were included as a mediator was significant, Sobel's $Z = 2.44$, $p = .02$.

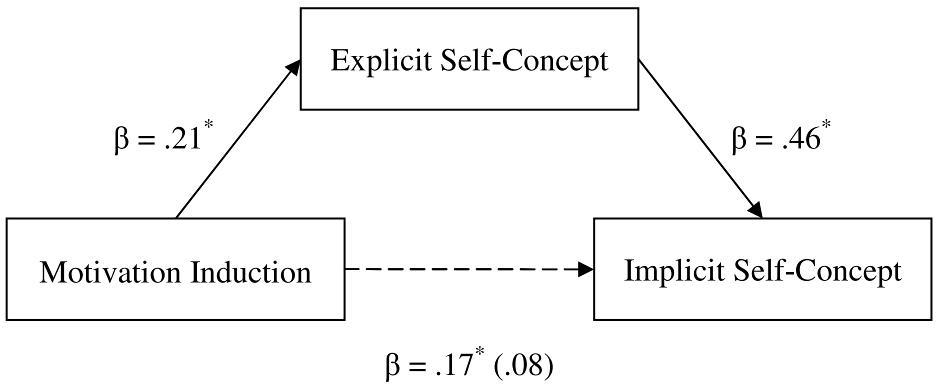

Figure 2.2. Mediation model tested in Experiment 2.2 (on the basis of Baron & Kenny, 1986). The indirect effect of the motivation induction on the implicit self-concept (mediated by the explicit self-concept) is statistically significant, Sobel's $Z = 2.44$, $p = .02$.
$^{*}p < .05$.

To rule out the alternative possibility of a bottom-up influence, we also tested the reverse mediation model, in which self-report scores were simultaneously regressed on both the motivation induction and the IAT scores. In this mediation model, changes in IAT scores failed to account for the obtained effect on self-report scores, in that the motivation induction still had a marginally significant effect on self-report scores after controlling for IAT scores, β = .14, $t(115) = 1.88$, $p = .06$. Thus, changes in the explicit self-concept fully accounted for changes in the implicit self-concept, but not the other way around.

2.2.3 Discussion

The results of Experiment 2 provided support for all three predictions concerning top-down self-construal. The motivation induction led to revision of the explicit, working self-concept in line with the desired trait and selectively activated congruent trait-related self-associations in the implicit self-concept. The mediation analysis supported the prediction that the effect of the motivation induction on the implicit self-concept would be mediated by changes in the explicit self-concept. Consistent with this prediction, changes in explicit self-beliefs fully accounted for changes in self-associations, but not the other way around. Taken together, these findings suggest that differences in participants' implicit self-concepts between motivation induction conditions were due to the top-down, intentional activation of self-associations in memory to substantiate the assertion of a propositional belief within the explicit, working self-concept.

2.3 General Discussion

The current experiments were designed to test a framework specifying how the explicit and implicit self-concepts are related through processes of mutual influence. The results of these two experiments provide converging evidence for the predictions derived from this framework regarding the roles of memory activation and motivated reasoning in achieving correspondence between these two aspects of self-representation. In Experiment 1, participants recalled specific autobiographical memories independent of the intention to revise the explicit, working self-concept. The recall task produced congruent changes in the implicit and explicit self-concepts and, consistent with the proposed account of bottom-up self-construal, changes in the implicit self-concept fully mediated changes in the explicit self-concept. In Experiment 2, participants were motivated to revise their explicit, working self-concepts directly by asserting the validity of a propositional self-belief. The induced motivation produced congruent changes in the explicit and implicit self-concepts; however, in this case—consistent with the proposed account of top-down self-construal—changes in the explicit self-concept fully mediated changes in the implicit self-concept. These results together support the claim that the explicit and implicit self-concepts are integrated, interacting aspects of a dynamic self-system.

Although the mediation analyses generally confirmed our predictions about bottom-up and top-down self-construal, a potential concern is that the reverse mediation models in both experiments revealed evidence for partial mediation (for similar findings, see Gawronski & Walther, 2008; Whitfield & Jordan, 2009). Specifically, the reverse mediation models showed simultaneous direct *and* indirect effects that were close to or at statistical significance.[9] These data patterns reflect an inherent limitation of correlation-based approaches to mediation, in which mediation is established on the basis of the shared covariance between two measured variables and an independent variable (cf. Zhao, Lynch, & Chen, 2010). The possibility of partial mediation, however, becomes theoretically implausible when examined alongside evidence from the predicted mediation models. In Experiment 1, the effect of the recall task on self-associations fully accounted for changes in explicit self-beliefs. Likewise, in Experiment 2, the effect of the motivation induction on explicit self-beliefs fully accounted for changes in self-associations. If the current manipulations influenced our dependent measures through processes of partial mediation, the proposed mediators in the two experiments would be unable to fully account for changes in the proposed distal outcomes. Rather, there should still be a direct effect on the distal outcome in the predicted mediation model after controlling for the proposed mediator. For instance, if the data in Experiment 1 reflected the operation of a direct influence on self-associations and a simultaneous indirect influence on self-associations mediated by a direct influence on explicit self-beliefs, the obtained effect on explicit self-beliefs should remain significant after controlling for self-associations. Similarly, if the data in Experiment 2 reflected the operation of a direct influence on self-beliefs and a simultaneous indirect influence on self-beliefs mediated by a direct influence on self-associations, the obtained effect on self-associations should remain significant after controlling for self-beliefs. This was not, however, the case. Instead, changes in self-associations fully accounted for the obtained effect on self-beliefs in Experiment 1, and changes in self-beliefs fully accounted for the obtained effect on self-associations in Experiment 2. These results are consistent with the current

---

[9] In Experiment 1, the reverse mediation model revealed a marginally significant direct effect, $\beta = .14$, $t(115) = 1.76$, $p = .08$, and a marginally significant indirect effect, Sobel's $Z = 1.89$, $p = .06$; in Experiment 2, the reverse mediation model revealed a marginally significant direct effect, $\beta = .14$, $t(115) = 1.88$, $p = .06$, and a significant indirect effect, Sobel's $Z = 1.99$, $p = .05$.

hypotheses of bottom-up and top-down construal, but they are inconsistent with the alternative possibility of partial mediation.[10]

2.3.1 Correspondence vs. Dissociation

The present research emphasized the correspondence between measures of the explicit and implicit self-concepts, in contrast to previous research that has emphasized their dissociation (Schnabel & Asendorpf, 2010). An obvious question for the proposed framework of self-construal, then, is how to account for such dissociations. To begin with, the results of the current experiments provide evidence for mutual influences between the explicit and implicit self-concepts via a *knowledge-activation* process (Förster & Liberman, 2007). All else being equal, the activation of self-knowledge, whether occurring via a process of bottom-up or top-down self-construal, should increase correspondence between the explicit and implicit self-concepts. This correspondence may break down during either of these processes, however, resulting in a dissociation. First, in the case of bottom-up self-construal, the influence of self-associations on explicit self-beliefs is likely moderated by a *belief-validation* process. Following Gawronski and Bodenhausen (2006, in press), the belief-validation process is expected to operate according to principles of cognitive consistency, such that activated self-associations that are inconsistent with other (subjectively valid) self-beliefs may be rejected as invalid information about the self. Whereas validation of activated self-associations should increase the correspondence between the explicit and implicit self-concepts, invalidation should result in a dissociation within the relevant domain of self-knowledge. Second, in the case of top-down self-construal, dissociations may arise when new beliefs are asserted as valid within the explicit, working self-concept, but are not substantiated via selective activation of confirmatory self-associations. Thus, whereas selective activation

---

[10] Further evidence for our mediation hypotheses could be obtained through experimental approaches that do not rely on simple covariations between the mediator and the distal outcome (Spencer, Zanna, & Fong, 2005). One option is to experimentally manipulate the effect of the proposed mediator on the distal outcome (see Gawronski & LeBel, 2008, for an example). To the extent that the effect of the mediator on the distal outcome can be disrupted, the effect of the original manipulation (e.g., the motivation induction) should remain intact for the mediator (e.g., explicit self-concept), but it should disappear for the distal outcome (e.g., implicit self-concept).

of confirmatory self-associations should increase the correspondence between the explicit and implicit self-concepts, disrupting the process of confirmatory information search should lead to a dissociation.

Accounting for both correspondence and dissociation between the explicit and implicit self-concepts suggests a more comprehensive framework for understanding self-construal as an epistemic enterprise, characterized in terms of the basic principles of knowledge-activation and belief-validation (Gawronski, LeBel, & Peters, 2007). Such a framework has the potential to clarify both how the explicit and implicit self-concepts correspond and how they become dissociated.

2.3.2 Future Directions

Based on the above discussion, an important next step in the development of this framework is to investigate the proposed account of self-concept dissociations. In particular, the framework predicts that the overall self-system comprising the explicit and implicit self-concepts can become "unbalanced" when the processes that maintain correspondence break down. On the one hand, inconsistent beliefs implied by self-associations activated within the implicit self-concept may not be validated for incorporation into the explicit, working self-concept (though self-associations may nevertheless influence spontaneous behaviors; Asendorpf et al., 2002). On the other hand, propositional beliefs asserted within the explicit, working self-concept may remain unsubstantiated if the activation of confirmatory self-associations in memory is interrupted. The resulting discrepancies may promote uncertainty in self-definition (Briñol et al., 2006) and compensatory behaviors intended to substantiate the asserted self-beliefs (e.g., Wicklund & Gollwitzer, 1981). Thus, the current framework not only offers specific predictions about the mutual influences between these two aspects of self-representation, but also integrates earlier findings on the dynamics of the explicit and implicit self-concepts, providing intriguing directions for future research.

2.4 References

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between explicit and implicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, *83*, 380-393.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*, 533-548.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Briñol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, *91*, 154-170.

Cross, S. E., & Markus, H. R. (1990). The willful self. *Personality and Social Psychology Bulletin*, *16*, 726-742.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. Wiers & A. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage.

Förster, J., & Liberman, N. (2007). Knowledge activation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 201-231). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731.

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

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355-1361.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2,* 181-193.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34,* 1276-1289.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022-1038.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197-216.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369-1385.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211-228.

Kruglanski, A. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

Libby, L. K., Eibach, R. P., & Gilovich, T. (2005). Here's looking at me: The effect of memory perspective on assessments of personal change. *Journal of Personality and Social Psychology*, *88*, 50-62.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224-253.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*, 299-337.

Quine, W. V. O., & Ullian, J. S. (1970). *The web of belief*. New York: Random House.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology*, *59*, 229-241.

Schnabel, K., & Asendorpf, J. B. (2010). The self-concept: New insights from implicit measurement procedures. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 408-425). New York: Guilford Press.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than meditational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845-851.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). San Diego, CA: Academic Press.

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology, 45,* 748-759.

Wicklund, R. A., & Gollwitzer, P. M. (1981). Symbolic self-completion, attempted influence, and self-deprecation. *Basic and Applied Social Psychology*, *2*, 89-114.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37,* 197-206.

2.5 Appendix A

Implicit Association Test Stimuli

The following tables list the stimuli used in the Implicit Association Test (IAT) for the target (“me” vs. “not me”) and attribute (“extravert” vs. “introvert”) categories in Experiments 1 and 2.

Target words		Attribute words	
“me”	“not me”	“extravert”	“introvert”
I	*few*	*active*	*passive*
me	*some*	*talkative*	*quiet*
my	*any*	*sociable*	*withdrawn*
mine	*it*	*outgoing*	*private*
self	*other*	*assertive*	*reserved*

2.6 Appendix B

Self-Report Scale of Self-Perceived Personality Traits

For each of the 24 items in the scale, subjects rated the statement "I am X," where X was one of the personality traits below, on a 7-point agree/disagree scale. The same scale was used in Experiments 1 and 2.

Extraversion	Introversion	Filler (positive)	Filler (negative)
active	*passive*	*curious*	*anxious*
talkative	*quiet*	*disciplined*	*impulsive*
sociable	*withdrawn*	*generous*	*selfish*
outgoing	*reserved*	*humorous*	*dishonest*
assertive	*private*	*optimistic*	*cynical*
extraverted	*introverted*	*rational*	*superstitious*

2.7 Appendix C

Manipulation of Motivation in Experiment 2

The following text was presented in the form of a fake newspaper clipping (adapted from Sanitioso et al., 1990). The text used in the Pro-Extravert condition is shown here; in the Pro-Introvert condition, all references to extraversion and introversion were switched.

> ***Extraverts get ahead***
>
> A recent study at Stanford University investigating the effects of extraverted personality on academic and job success has concluded that outgoing people are more successful than their less talkative peers. Dr. Brian Carswell, the lead investigator, reports that although roughly equal numbers of extraverts and introverts exist in the population, being extraverted appears to predict success in these settings to a high degree. In particular, Carswell and his colleagues found that extraverts tend to receive higher grades in school and are more likely to earn graduate and professional degrees compared to introverts. Carswell also reports that extraverted individuals are more likely to end up in successful, high-paying careers. “Extraversion appears to confer distinct advantages in the modern world,” Carswell said.

3 Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations

In the run-up to the 2008 US Presidential election, as the competition between Barack Obama and John McCain intensified, *The New Yorker* magazine published a cover illustration depicting Obama as a terrorist occupying the Oval Office. The resulting uproar from Obama's supporters reflected a suspicion that the illustration could lead voters to form negative associations with Obama even if they rejected the depicted link to terrorism (Banaji, 2008). Negative political campaigning has likewise been criticized for exploiting the ease with which evaluative associations can be manipulated (e.g., Carraro, Gawronski, & Castelli, 2010), and the veracity of these associations matters, as they have been shown to predict voting behavior in undecided voters independent of conscious beliefs (Galdi, Arcuri, & Gawronski, 2008; Payne, Krosnick, Pasek, Lelkes, Akhtar, & Thompson, 2010). These examples suggest that learning might involve more than just the formation of beliefs, such that evaluative associations might be formed independent of, and even despite, conscious assessment of their validity. The unnerving implication is that we may be no more than "puppets on a string," helpless to resist being influenced by all the contingencies to which we are exposed in an information-saturated world.

The difference between associations and beliefs is supported by research in social cognition showing that it is possible for people to express divergent evaluations of the same object under different conditions (see Gawronski & Bodenhausen, 2006, for a review). For instance, explicit evaluations expressed under conditions of controlled processing often diverge from implicit evaluations expressed under conditions of automatic processing (Bargh, 1994). These results suggest that evaluative responses do not always reflect conscious beliefs about an object and that automatically activated associations may influence evaluative responses under suboptimal processing conditions.

In line with suspicions about the effects of media influence, evaluative dissociations are often explained by appeal to dual-process theories of learning, which posit two learning processes that may operate in parallel (e.g., Gawronski & Bodenhausen, 2006; Rydell &

McConnell, 2006; Strack & Deutsch, 2004). On this account, explicit evaluations are the product of a *belief-based learning* process, which qualifies the evaluation implied by object-valence contingencies by the perceived validity of these contingencies. Implicit evaluations, in contrast, are thought to be the product of a *contingency-based learning* process, which encodes contingencies independent of their perceived validity and hence is sensitive to the dominant valence associated with an object. Because the two learning processes are assumed to operate in parallel, evaluative dissociations may already occur at the time of learning if the evaluation implied by observed contingencies is qualified by conscious beliefs about its validity.

Indeed, evidence for dual-process theories of learning seems compelling, such that evaluative dissociations may often arise at the time of learning about an object. For example, Rydell, McConnell, Mackie, and Strain (2006) employed a learning procedure in which participants had to guess the validity of valenced behavioral descriptions about a social target named Bob. Preceding the display of Bob's photograph on each trial, participants were subliminally presented with a prime word whose valence was opposite to the evaluation implied by the validity of the behavioral descriptions. Rydell et al. found that explicit evaluations of Bob reflected the valence of the valid behavioral descriptions whereas implicit evaluations of Bob reflected the valence of the subliminal primes.

Although findings like these provide evidence that implicit and explicit evaluations are differentially sensitive to contingency-based versus belief-based learning processes, they remain silent about whether the two learning mechanisms can operate *simultaneously* on the basis of the *same* information (as is implied in the depiction of Obama as a terrorist). In particular, the currently available evidence is limited in answering this question given that demonstrations of learning-related dissociations typically involved multiple manipulations of an object's valence via distinct sources of information. For instance, in Rydell et al.'s (2006) study, contingency-based learning was driven by the subliminal primes that preceded the presentation of Bob, whereas belief-based learning was driven by the validity of the behavioral descriptions that followed the presentation of Bob. Moreover, the contingencies established through the priming manipulation may not be subject to conscious qualification if they are learned outside awareness. Consequently, it remains unclear if evaluative dissociations can arise in situations in which the evaluation implied by contingencies can itself be assessed as true or false. Such conditions are more consistent with the examples of media influence described above, where the observer is frequently confronted with information that is immediately perceived to be invalid.

The present experiments aim to address this question by directly manipulating the perceived validity of object-valence contingencies in a single learning episode. Based on dual-process theories that propose two parallel learning mechanisms (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004), we expected an evaluative dissociation in this situation due to the simultaneous operation of belief-based and contingency-based learning processes. Specifically, given evidence that the negation of an association has been shown to qualify explicit, but not implicit, evaluations (Deutsch, Gawronski, & Strack, 2006; Gregg, Seibt, & Banaji, 2006), evaluative dissociations were expected to arise when the valence most frequently associated with an object is perceived to be false. In these cases, explicit evaluations should reflect the perceived validity of observed contingencies, whereas implicit evaluations should reflect observed contingencies independent of their perceived validity.

To anticipate our findings, this prediction was *not* confirmed in our studies. Contrary to the assumption that contingency-based and belief-based learning mechanisms may operate simultaneously on the basis of the same information, we found that the perceived validity of evaluative information about social targets qualified both explicit and implicit evaluations when validity information was available during the learning of the valence information. The expected dissociations only occurred when the presentation of validity information was delayed, which reduced its qualifying effect on implicit, but not explicit, evaluations. Taken together, these results support accounts that explain evaluative dissociations in terms of *expression-related*, as opposed to *learning-related* processes, such that dissociations can be explained by the rejection of previously formed associations during the course of generating a controlled evaluative response (see Hofmann, Gschwendner, Nosek, & Schmitt, 2005, for a review). Consequently, these findings pose a challenge to the view that evaluative dissociations may be explained by the simultaneous operation of two independent learning mechanisms on the basis of the same information.

3.1 Experiment 1

In the first experiment, we sought to test the basic question of whether an evaluative dissociation can arise during a single learning episode, in which the perceived validity of object-valence contingencies is directly manipulated. The experiment involved a social learning task that required participants to form impressions of four novel targets by reading valenced behavioral descriptions about each of them. To manipulate the perceived validity of the resulting contingencies, each behavioral description was immediately marked as either true or false. The four targets in the learning task thus differed according to the valence with which they were most frequently associated (positive vs. negative) as well as the "true" valence of each target. The key empirical question is whether an evaluative dissociation will arise when the "true" valence diverges from the valence most frequently associated with the target. Based on the assumption that contingency-based and belief-based learning mechanisms may operate simultaneously on the basis of the same information, we originally expected that explicit evaluations would reflect the observed contingencies qualified by their perceived validity, whereas implicit evaluations would reflect these contingencies without any qualification.

3.1.1 Method

3.1.1.1 Participants and design

A total of 28 undergraduate students (20 women; 8 men) participated in a study on impression formation for course credit. The experiment employed a 2 (Dominant Valence: 75% Positive vs. 75% Negative) × 2 (Validity of Dominant Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) factorial design with all three variables varying within-participants. Order of the two evaluation measures was counterbalanced across participants.

3.1.1.2 Learning task

Upon entering the lab, participants were seated at individual computers and signed informed consent documents. Participants then began the computerized learning task, in which they were asked to form accurate impressions of four different people based on minimal information about each of them. The learning paradigm consisted of a guessing task in which participants were sequentially presented with photographs of four men along with valenced behavioral descriptions of each of them. Participants' task was to guess the accuracy of each description with feedback provided immediately following each guess (i.e., "RIGHT!" or "WRONG!"; see Rydell et al., 2006). Feedback following guesses was 100% consistent for all targets so that there was no misleading feedback about the true valence of any target. Each participant thus learned the true valence of each target according to his or her own pattern of guesses. In addition, instructions preceding the task informed participants that, when a behavioral description turned out to be false, they should infer that the opposite of the implied evaluation was true (i.e., a positive description that turns out to be false implies a negative evaluation, and vice versa).

A total of 20 behavioral descriptions (adapted from Rydell et al., 2006) were presented for each of the four targets, including both positively valenced descriptions (e.g., "Mike lent money to a friend in financial trouble") and negatively valenced descriptions (e.g., "Mike cheated during a poker game"). The valence of the behavioral descriptions for each target was held in 3:1 proportion such that two targets were paired with 15 positive and 5 negative descriptions and two targets were paired with 5 positive and 15 negative descriptions. The "true" valence was varied orthogonally to the dominant valence of the behavioral descriptions so that either the positive descriptions were true and the negative descriptions were false, or vice versa. These two manipulations created four different impression-formation targets: (1) a target with mostly positive descriptions that were described as accurate, (2) a target with mostly negative descriptions that were described as accurate, (3) a target with mostly positive descriptions that were described as inaccurate, and (4) a target with mostly negative descriptions that were described as inaccurate. The particular mappings of the four photographs with the four experimental conditions were counterbalanced across participants.

With 20 behavioral descriptions presented for each of the four targets, the learning procedure consisted of a total of 80 trials presented to each participant in computer-randomized order. Each learning trial started with the presentation of a shoulder-up photograph of one of the four impression-formation targets, all of whom were young, white men, centered on the computer screen. At the same time, a valenced behavioral description was displayed below the picture. Participants were instructed to use two response keys on the keyboard to indicate their true/false guess on each trial. Upon making a response, the display was cleared and participants were given feedback about the validity of their guess. The feedback remained centered on the screen for 1000 milliseconds, followed by a 1000 millisecond inter-trial interval.

3.1.1.3 Measurement of explicit evaluations

Following the learning procedure, participants completed measures of explicit and implicit evaluations in counterbalanced order. To assess explicit evaluations, participants completed three self-report items (likeability, friendliness, and trustworthiness) for each of the four impression-formation targets in computer-randomized order. Responses for all items were made on 7-point rating scales ranging from 1 (*not at all*) to 7 (*very much*).

3.1.1.4 Measurement of implicit evaluations

The affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) was used to measure implicit evaluations of each of the four impression-formation targets. Each trial of the AMP was displayed in the following sequence: A fixation cross was presented for 1000 milliseconds; a valenced stimulus (i.e., a photograph of one of the four targets) for 75 milliseconds; a blank screen for 125 milliseconds; a Chinese pictograph for 100 milliseconds; and finally, a pattern mask of black and white noise was presented. Participants were instructed that, upon presentation of the mask, they were to indicate how “visually pleasant” they found the preceding Chinese pictograph using two response keys on the keyboard signifying *less pleasant* and *more pleasant*. Following Payne et al. (2005), participants were told that the pictures appearing before the Chinese pictographs may bias responses and that they should try not to let the pictures influence their judgments. Twenty-five AMP trials were presented for each impression-formation target, resulting in a total of 100 trials presented in computer-randomized order. Participants were debriefed following completion of the dependent measures.

3.1.2 Results

3.1.2.1 Data preparation

The three self-report items were averaged to create an index of the explicit evaluation of each of the four targets (all Cronbach's $\alpha > .56$). To create an index of the implicit evaluation of each target, the proportion of *more pleasant* responses on the relevant AMP trials was calculated, which varied between 0% (negative) and 100% (positive).

3.1.2.2 Explicit and implicit evaluations

The primary analyses collapsed across the order of the two evaluation measures. To test the effects of validity information on explicit and implicit evaluations, indices of both explicit and implicit evaluations were standardized to obtain a common metric and then submitted to a 2 (Dominant Valence: Positive vs. Negative) × 2 (Validity of Dominant Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) repeated measures analysis of variance (ANOVA). Significant main effects were observed for valence, $F(1, 27) = 19.07$, $p < .001$, $\eta^2_p = .41$, and validity, $F(1, 27) = 11.77$, $p = .002$, $\eta^2_p = .30$. In addition, significant two-way interactions were observed between valence and validity, $F(1, 27) = 73.07$, $p < .001$, $\eta^2_p = .73$, and validity and evaluation type, $F(1, 27) = 25.89$, $p < .001$, $\eta^2_p = .49$. Finally, the three-way interaction between valence, validity, and evaluation type was significant, $F(1, 27) = 17.63$, $p < .001$, $\eta^2_p = .40$. Further inspection of this interaction suggests that the qualification of the validity × valence effect by evaluation type does not reflect the expected dissociation between explicit and implicit evaluations as a function of validity. Instead, the interaction simply reflects a slightly weaker effect size of the valence × validity cross-over interaction for implicit evaluations, as described below.

To specify the obtained three-way interaction, the effects of the valence and validity manipulations were assessed separately for both explicit and implicit evaluations using raw scores for all analyses. With respect to explicit evaluations, significant main effects of valence, $F(1, 27) = 22.53$, $p < .001$, $\eta^2_p = .46$, and validity, $F(1, 27) = 64.57$, $p < .001$, $\eta^2_p = .63$, were observed, qualified by a significant two-way interaction, $F(1, 27) = 72.30$, $p < .001$, $\eta^2_p = .73$. As shown in Figure 3.1, validity information influenced explicit evaluations as expected, such that explicit evaluations reflected the dominant valence when it was true and the opposite of the dominant valence when it was false. Paired-samples *t*-tests revealed that when the dominant valence was true, explicit evaluations favored positively described targets over negatively described targets, $t(27) = 9.08$, $p < .001$; but when the dominant valence was false, explicit evaluations favored negatively described targets over positively described targets, $t(27) = 6.04$, $p < .001$. Moreover, when the dominant valence of behavioral descriptions was positive, explicit evaluations were more positive when the validity feedback for the dominant information was true rather than false, $t(27) = 12.80$, $p < .001$; but when the dominant valence of behavioral descriptions was negative, explicit evaluations were more positive when the validity feedback for the dominant information was false rather than true, $t(27) = 4.28$, $p < .001$.

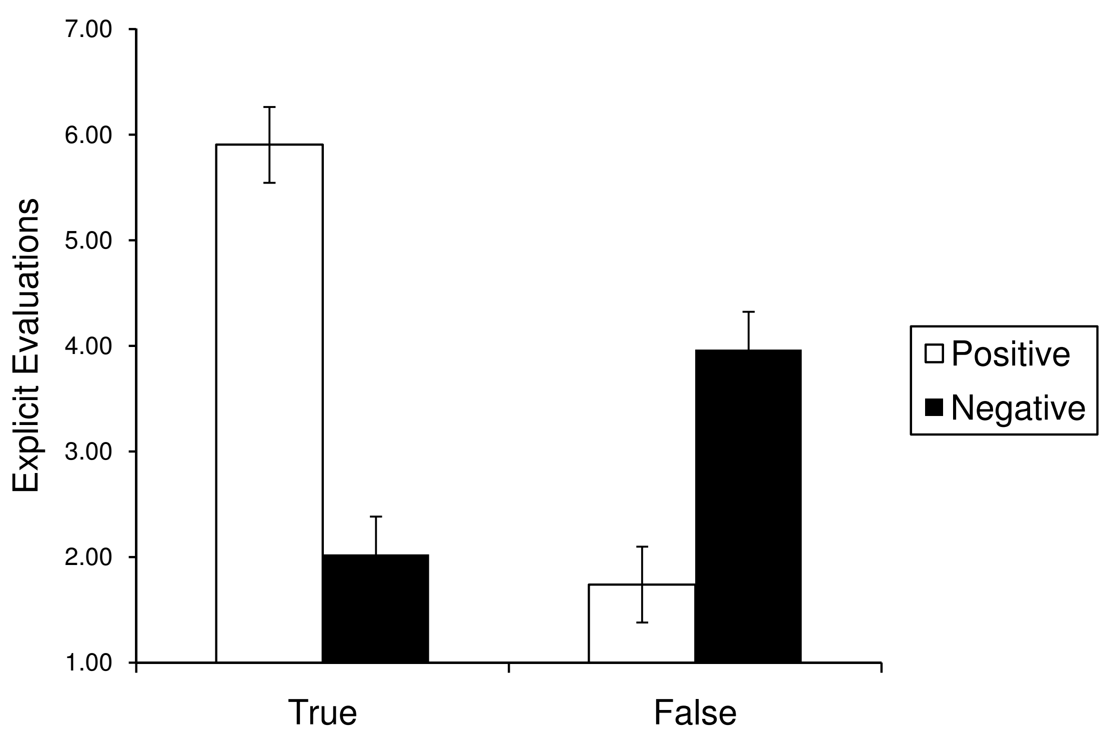

Figure 3.1. Explicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false), Experiment 3.1. Error bars represent standard errors.

With respect to implicit evaluations, no main effects were significant but the two-way interaction between valence and validity was significant, $F(1, 27) = 16.85$, $p < .001$, $\eta^2_p = .38$. As shown in Figure 3.2, the interaction effect for implicit evaluations was identical to that obtained for explicit evaluations; that is, implicit evaluations reflected the dominant valence when it was true and the opposite of the dominant valence when it was false. Paired-samples *t*-tests revealed that when the dominant valence was true, implicit evaluations favored positively described targets over negatively described targets, $t(27) = 4.03$, $p < .001$; but when the dominant valence was false, implicit evaluations favored negatively described targets over positively described targets, $t(27) = 2.66$, $p = .013$. Moreover, when the dominant valence of behavioral descriptions was positive, implicit evaluations were more positive when the validity feedback for the dominant information was true rather than false, $t(27) = 3.16$, $p = .004$; but when the dominant valence of behavioral descriptions was negative, implicit evaluations were more positive when the validity feedback for the dominant information was false rather than true, $t(27) = 3.44$, $p = .002$.[11]

---

[11] The order of the two evaluation measures did not moderate the effect of valence and validity information on the measure of implicit evaluations, $F(1, 26) = .02$, $p = .879$, $\eta^2_p < .01$, but did moderate the effect of valence and validity on the measure of explicit evaluations, $F(1, 26) = 6.59$, $p = .016$, $\eta^2_p = .20$. The order effect reflects a stronger effect of the valence × validity interaction on explicit evaluations when the measure of explicit evaluations was completed first, although in both cases the interaction remained significant. Specifically, when the measure of explicit evaluations was completed after the measure of implicit evaluations, the two-way interaction between valence and validity was relatively weaker, $F(1, 13) = 15.62$, $p = .002$, $\eta^2_p = .55$, compared to when it was completed first, $F(1, 13) = 133.60$, $p < .001$, $\eta^2_p = .91$. Because measurement order had no effect on the results in Experiments 2 and 3, we refrain from speculating on the nature of this effect.

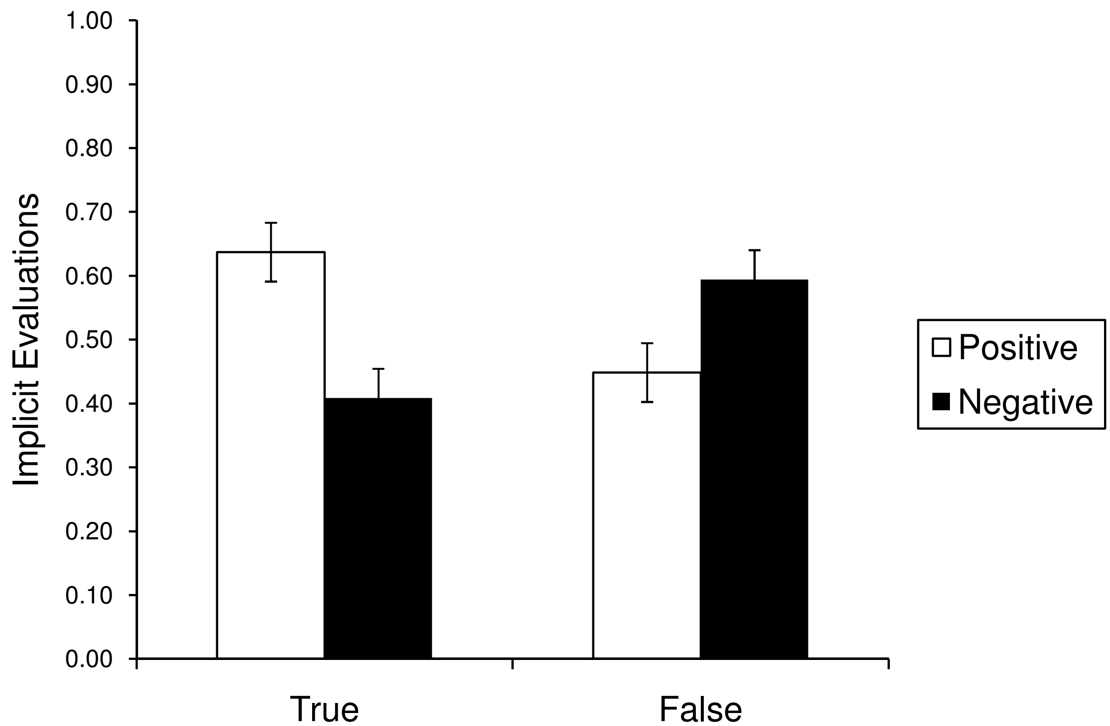

Figure 3.2. Implicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false), Experiment 3.1. Error bars represent standard errors.

3.1.3 Discussion

Counter to our predictions, Experiment 1 revealed that perceived validity qualified the effect of object-valence contingencies for both explicit and implicit evaluations. When the dominant valence was true, explicit and implicit evaluations reflected the dominant valence, but when the dominant valence was false, explicit and implicit evaluations reflected the opposite of the dominant valence. Thus, no evaluative dissociation was observed when the perceived validity of observed contingencies was manipulated in a single learning episode. This pattern of results challenges the assumption that belief-based and contingency-based learning may operate simultaneously on the basis of the same information. Drawing on dual-process theories that propose the simultaneous operation of two parallel learning mechanisms, we originally expected that explicit evaluations would reflect the perceived validity of contingencies, whereas implicit evaluations would reflect contingencies independent of their perceived validity. This prediction was clearly disconfirmed in the current study. There is, however, a methodological concern with drawing this conclusion directly from the present results, which was addressed in the next experiment.

3.2 Experiment 2

The present experiment sought to rule out the concern that the absence of a dissociation in Experiment 1 resulted from inadequate measurement procedures. Recent evidence suggests that validity information pertaining to the primes can influence responses on the AMP (Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009), whereas Fazio's evaluative priming task (EPT; Fazio, Jackson, Dunton, & Williams, 1995) remains unaffected by validity information (Deutsch et al., 2006, 2009). Moreover, the AMP and the EPT have been shown to produce divergent effects of the same experimental manipulation under some conditions, suggesting that task-specific mechanisms may shape responses on these measures in a non-trivial manner (e.g., Deutsch & Gawronski, 2009; Gawronski, Cunningham, LeBel, & Deutsch, 2010). It would therefore be valuable to replicate the findings from Experiment 1 using an EPT to ensure that they are not unique to one measure of implicit evaluations but reflect a genuine effect that replicates across different measures of the same construct.

3.2.1 Method

3.2.1.1 Participants and design

A total of 45 undergraduate students (37 women; 8 men) participated in a study on impression formation for course credit. One participant was excluded from analysis due to chance responding on the EPT (error rate > 40%). As with Experiment 1, this experiment employed a 2 (Dominant Valence: 75% Positive vs. 75% Negative) × 2 (Validity of Dominant Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) factorial design with all three variables varying within-participants. Order of the two evaluation measures was counterbalanced across participants.

3.2.1.2 Procedure

The learning task and the measure of explicit evaluations were identical to Experiment 1. Fazio et al.'s (1995) EPT was used to measure implicit evaluations of the four impression-formation targets. Each trial of the EPT was displayed in the following sequence: A fixation cross was presented for 500 milliseconds; a valenced prime (i.e., a photograph of one of the four targets) for 200 milliseconds; and then a positive or negative target word (e.g., "paradise" or "poison"), which remained onscreen until the participant indicated whether the word was positive or negative using one of two response keys on the keyboard. If the response was incorrect, "*ERROR!*" was displayed for 1500 milliseconds. An interval of 1000 milliseconds preceded the start of the next trial. According to Fazio et al. (1995), the affect elicited by the prime should facilitate evaluative decisions for valence-congruent target words but inhibit evaluative decisions for valence-incongruent target words, so that response latencies to the target words can be used to infer implicit evaluations of each impression-formation target. Each of the four targets served as a prime on 20 trials, split between 10 trials with negative and 10 trials with positive target words, creating a total of 80 trials presented in computer-randomized order. Participants were debriefed following completion of the dependent measures.

3.2.2 Results

3.2.2.1 Data preparation

Indices of the explicit evaluation of each of the four impression-formation targets were calculated as described in Experiment 1 (all Cronbach's $\alpha > .83$). In creating indices of the implicit evaluation of each of the four impression-formation targets, EPT trials with incorrect responses (5.1%) were excluded. To control for anticipations and outliers (Ratcliff, 1993), response cutoffs were employed to exclude trials with reaction times shorter than 300 milliseconds or longer than 1000 milliseconds (8.7% of valid trials). Then, for each of the four primes in the EPT, the mean reaction time to trials with positive target words was subtracted from trials with negative target words, so that higher scores reflect a relatively more positive implicit evaluation of the target (see Wentura & Degner, 2010).

3.2.2.2 Explicit and implicit evaluations

The order of the two evaluation measures had no effect, so analyses collapsed across this factor. To test the effects of validity information on explicit and implicit evaluations, indices of both explicit and implicit evaluations were standardized and submitted to a 2 (Dominant Valence: 75% Positive vs. 75% Negative) × 2 (Validity of Dominant Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) repeated measures ANOVA. Significant two-way interactions were observed between valence and validity, $F(1, 43) = 358.08$, $p < .001$, $\eta^2_p = .89$, and valence and evaluation type, $F(1, 43) = 4.53$, $p = .039$, $\eta^2_p = .10$. The three-way interaction between valence, validity, and evaluation type was also significant, $F(1, 43) = 270.53$, $p < .001$, $\eta^2_p = .86$. No other effects were significant. As with Experiment 1, the qualification of the validity × valence effect by evaluation type does not reflect the expected dissociation between explicit and implicit evaluations. Instead, the interaction reflects a slightly weaker effect size of the valence × validity cross-over interaction for implicit evaluations, as described below.

To specify the obtained three-way interaction, the effects of the valence and validity manipulations were assessed separately for both explicit and implicit evaluations, using raw scores for all analyses. With respect to explicit evaluations, significant main effects were observed for valence, $F(1, 43) = 8.10$, $p = .007$, $\eta^2_p = .16$, and for validity, $F(1, 43) = 5.39$, $p = .025$, $\eta^2_p = .11$, qualified by a significant two-way interaction, $F(1, 43) = 445.49$, $p < .001$, $\eta^2_p = .91$. As shown in Figure 3.3, validity information influenced explicit evaluations as expected, such that explicit evaluations reflected the dominant valence when it was true and the opposite of the dominant valence when it was false. Paired-samples *t*-tests revealed that when the dominant valence was true, explicit evaluations favored positively described targets over negatively described targets, $t(43) = 21.60$, $p < .001$; but when the dominant valence was false, explicit evaluations favored negatively described targets over positively described targets, $t(43) = 16.67$, $p < .001$. Moreover, when the dominant valence of behavioral descriptions was positive, explicit evaluations were more positive when the validity feedback for the dominant information was true rather than false, $t(43) = 21.98$, $p < .001$; but when the dominant valence of behavioral descriptions was negative, explicit evaluations were more positive when the validity feedback for the dominant information was false rather than true, $t(43) = 17.17$, $p < .001$.

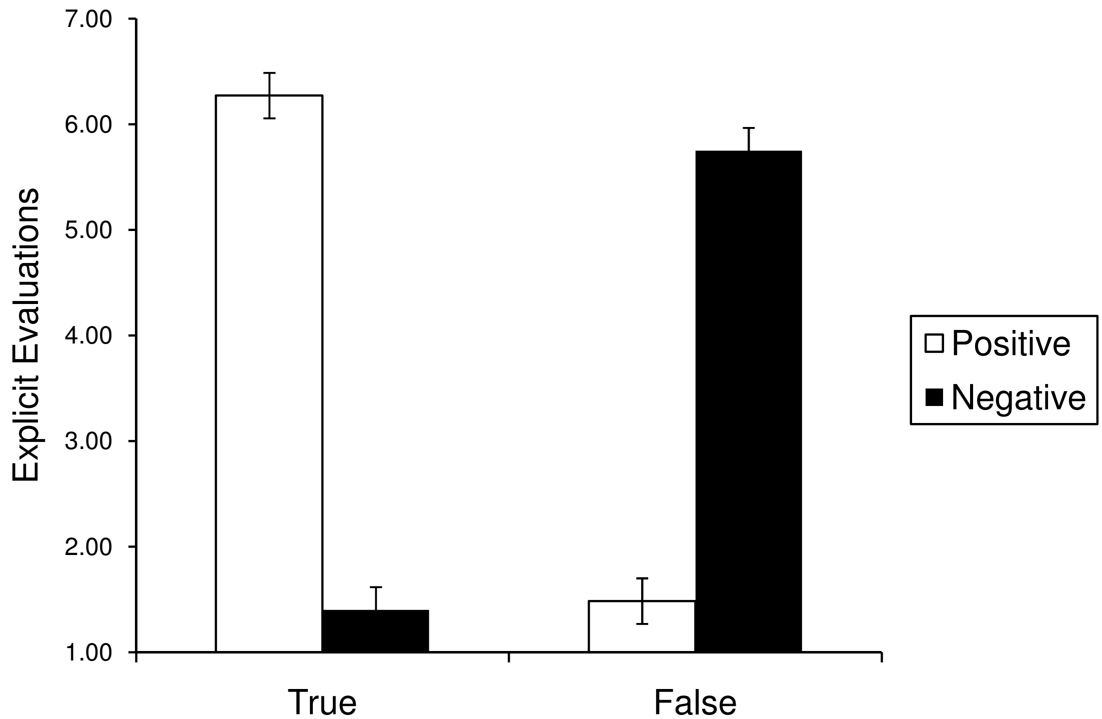

Figure 3.3. Explicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false), Experiment 3.2. Error bars represent standard errors.

With respect to implicit evaluations, no main effects were significant but the two-way interaction between valence and validity was significant, $F(1, 43) = 8.31$, $p = .006$, $\eta^2_p = .16$. As shown in Figure 3.4, the effect of validity information on implicit evaluations was identical to its effect of explicit evaluations; that is, implicit evaluations reflected the dominant valence when it was true and the opposite of the dominant valence when it was false. Paired-samples *t*-tests revealed that when the dominant valence was true, implicit evaluations favored positively described targets over negatively described targets, $t(43) = 2.15$, $p = .037$; but when the dominant valence was false, implicit evaluations favored negatively described targets over positively described targets, $t(43) = 2.29$, $p = .027$. Moreover, when the dominant valence of behavioral descriptions was positive, implicit evaluations were more positive when the validity feedback for the dominant information was true rather than false, $t(43) = 2.11$, $p = .041$; but when the dominant valence of behavioral descriptions was negative, implicit evaluations were more positive when the validity feedback for the dominant information was false rather than true, $t(43) = 2.42$, $p = .020$.

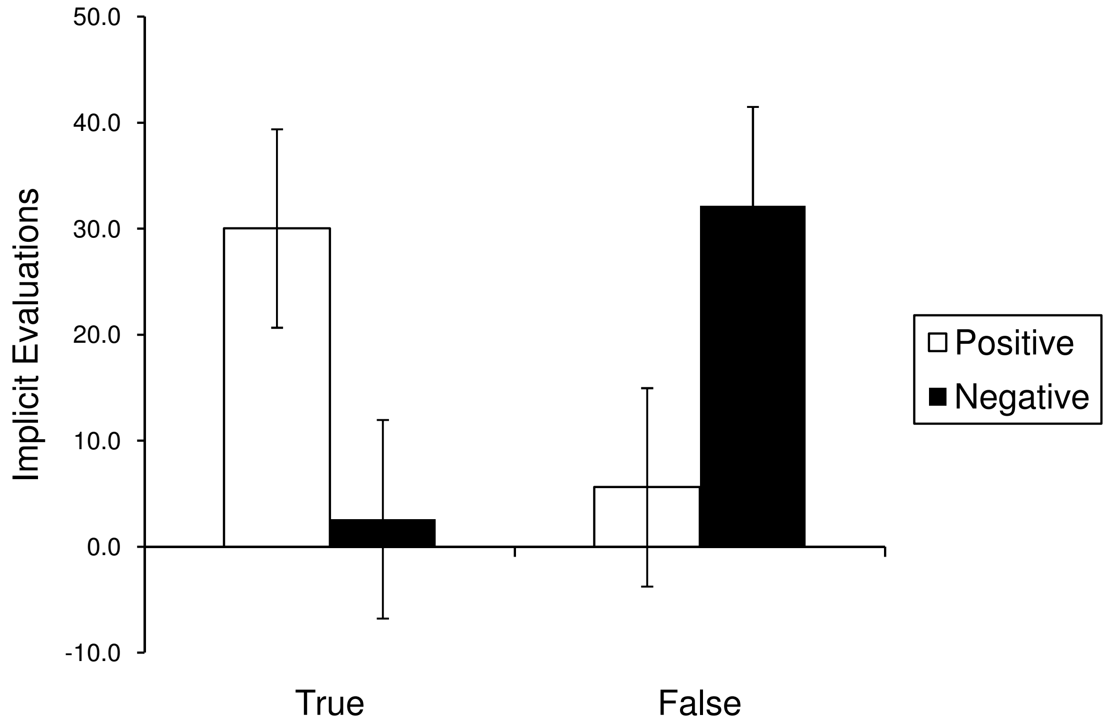

Figure 3.4. Implicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false), Experiment 3.2. Error bars represent standard errors.

3.2.3 Discussion

The findings of Experiment 2 replicated those of Experiment 1 using a measure of implicit evaluations less sensitive to validity information (Deutsch et al., 2006, 2009). Once again, perceived validity qualified the effect of object-valence contingencies for both explicit and implicit evaluations. These results rule out the concern that the absence of dissociation in Experiment 1 was due to suboptimal measurement procedures.

3.3 Experiment 3

Experiments 1 and 2 suggest that during a single learning episode, the perceived validity of object-valence contingencies influences both explicit and implicit evaluations. In other words, it appears that in situations that involve exposure to information that is considered invalid, it is possible to exercise control over what is learned. There is, however, evidence for evaluative dissociations arising from asymmetric influences of validity information on explicit and implicit evaluations. For example, Gregg et al. (2006) found that both explicit and implicit evaluations initially reflected the valence of behavioral descriptions of two novel groups, but that subsequently acquired information about the validity of these descriptions qualified explicit, but not implicit, evaluations. This finding is at odds with the results of the preceding experiments, in which validity information qualified both explicit and implicit evaluations.

An important factor that may account for the difference between the two findings is the time at which validity information was provided. Whereas in our studies validity information was available during the learning of the behavioral descriptions, Gregg et al.'s (2006) study included a substantial delay between the initial learning of the behavioral descriptions and the subsequent presentation of validity information. Thus, counter to the notion of *learning-related* dissociations due to the simultaneous operation of two distinct learning mechanisms on the basis of the same information, Gregg et al.'s (2006) findings are better described as a case of *expression-related* dissociations. Such dissociations occur when information that has been stored in memory at an earlier time is later learned to be invalid. In such cases, newly acquired validity information may be unable to erase previously formed associations from memory, even though these associations are rejected as invalid in the course of expressing an explicit evaluative judgment (Fazio, 2007; Gawronski & Bodenhausen, 2006). As a result, newly acquired validity information will influence explicit, but not implicit, evaluations. In fact, many examples of evaluative dissociations can be parsimoniously explained by the subsequent rejection of previously learned information without assuming a simultaneous operation of two independent learning mechanisms (see Hofmann et al., 2005, for a review).

If this interpretation is correct, then it should be possible to create an evaluative dissociation using the present experimental paradigm by manipulating the delay between the presentation of the behavioral descriptions and information about their validity. Experiments 1 and 2 showed that, when validity information is available during the learning of the behavioral descriptions, perceived validity produces equivalent effects on explicit and implicit evaluations. On the other hand, if the presentation of validity information is delayed, it must be applied to existing associations post-hoc, presumably qualifying explicit, but not implicit, evaluations. This pattern would be consistent with the idea that evaluative dissociations arising from perceived validity are due to expression-related, rather than learning-related, processes. Experiment 3 was designed to test this hypothesis.

3.3.1 Method

3.3.1.1 Participants and design

A total of 218 undergraduate students (159 women and 59 men; mean age = 22.03) participated in a study on impression formation for course credit. Data from 14 participants were unusable due to a programming error and another 15 participants were excluded due to chance responding on the EPT (error rates > 40%). The final sample consisted of 189 students (139 women; 50 men). The experiment employed a 2 (Valence: Positive vs. Negative) × 2 (Validity of Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) × 2 (Validity Timing: Short-Delay vs. Long-Delay) factorial design with the first three variables varying within-participants and the last varying between-participants. Order of the two evaluation measures was counterbalanced across participants.

3.3.1.2 Learning procedure

The learning procedure employed in Experiment 3 was broadly similar to that used in Experiments 1 and 2, with a few important differences. First, a more detailed cover story was provided, which framed the learning procedure in terms of learning about co-workers at a new job based on second-hand comments (adapted from Gawronski & Walther, 2008). To strengthen the overall effect of valence during the learning procedure, the four impression-formation targets were paired with 100% positive or 100% negative behavioral descriptions. The guessing component of the procedure was therefore dropped and the learning task was instead introduced as a slideshow that required only that participants attend to the information presented. Furthermore, because 100% consistent behavioral descriptions should be easily learned, only five learning trials were displayed for each target, for a total of 20 learning trials presented in computer-randomized order. New positive and negative behavioral descriptions were created to conform with the "workplace" cover story (adapted from Gawronski, Walther, & Blank, 2005).

To test the effects of immediate versus delayed presentation of validity information on explicit and implicit evaluations, the delivery of validity information during the learning procedure was manipulated to be either (a) interleaved with the learning trials in the short-delay condition or (b) presented after all learning trials had finished in the long-delay condition. Instructions prior to the learning task in the short-delay condition informed participants that some behavioral descriptions would turn out to be false and that in these cases they should infer that the opposite of the implied evaluation was true. In the long-delay condition, instructions prior to the learning task informed participants that some behavioral descriptions would turn out to be false but that they should initially assume that all of the descriptions are true.

Each learning trial in the short-delay condition (similar to Experiments 1 and 2) began with the presentation of a photograph of one of the four targets together with a valenced behavioral description. After 3000 milliseconds, validity information was presented just below the behavioral description and remained onscreen for another 3000 milliseconds. A 1500 millisecond inter-trial interval preceded the start of the next learning trial. In the long-delay condition, each learning trial began with the presentation of a photograph of one of the four targets together with a valenced behavioral description. This information remained onscreen for 6000 milliseconds and a 1500 millisecond inter-trial interval preceded the start of the next trial. The total duration of the 20-trial slideshow in both conditions was 150 seconds.

Following completion of the slideshow in the short-delay condition, participants were asked to take a moment to integrate the behavioral descriptions with the validity information to arrive at a clear impression of each target and to proceed to the next component of the study at their own pace. In the long-delay condition, participants were told that the behavioral descriptions for two of the targets were all true whereas the behavioral descriptions for the other two targets were all false. A photograph of each target and the validity of the descriptions associated with that target (i.e., "TRUE COMMENTS" or "FALSE COMMENTS") were displayed on one screen to make this clear. Participants were asked to take their time to arrive at a clear impression of each target in light of the new validity information. In both conditions, the valence and validity of the four targets were crossed to produce a positive/true, positive/false, negative/true, and negative/false target. The particular mappings of the four photographs with the four experimental conditions were counterbalanced across participants.

3.3.1.3 Measurement of explicit and implicit evaluations

The measures of explicit evaluations of each target were identical to those used in Experiments 1 and 2. Implicit evaluations of each target were assessed using an EPT identical to that used in Experiment 2, except that the total number of trials was doubled to 160. Participants were debriefed following completion of the dependent measures.

3.3.2 Results

3.3.2.1 Data preparation

Indices of the explicit evaluation of each of the four impression-formation targets were calculated as described in Experiment 1 (all Cronbach's $\alpha > .90$). In creating indices of the implicit evaluation of each of the four impression-formation targets, EPT trials with incorrect responses (4.3%) were excluded. Response cutoffs were also employed to exclude trials with reaction times shorter than 300 milliseconds or longer than 1000 milliseconds (7.3% of valid trials). Calculation of the implicit indices from the EPT scores followed the procedure described in Experiment 2.

3.3.2.2 Explicit and implicit evaluations

The order of the two evaluation measures had no effect, so analyses collapsed across this factor. To test the effects of delayed validity information on explicit and implicit evaluations, indices of both explicit and implicit evaluations were standardized and submitted to a 2 (Valence: Positive vs. Negative) × 2 (Validity of Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) × 2 (Validity Timing: Short-Delay vs. Long-Delay) mixed-model ANOVA with repeated measures on the first three factors. Significant main effects were observed for valence, $F(1, 187) = 54.48$, $p < .001$, $\eta^2_p = .23$, and validity, $F(1, 187) = 13.28$, $p < .001$, $\eta^2_p = .07$. In addition, significant two-way interactions were observed between valence and validity, $F(1, 187) = 618.97$, $p < .001$, $\eta^2_p = .77$, between valence and timing, $F(1, 187) = 11.61$, $p = .001$, $\eta^2_p = .06$, between valence and evaluation type, $F(1, 187) = 11.58$, $p = .001$, $\eta^2_p = .06$, and between validity and evaluation type, $F(1, 187) = 4.64$, $p = .033$, $\eta^2_p = .02$. Significant three-way interactions were observed between valence, validity, and timing, $F(1, 187) = 18.91$, $p < .001$, $\eta^2_p = .09$, between valence, validity, and evaluation type, $F(1, 187) = 532.18$, $p < .001$, $\eta^2_p = .74$, and between validity, timing, and evaluation type, $F(1, 187) = 7.27$, $p = .008$, $\eta^2_p = .04$. Finally, and most relevant to the current hypothesis, a significant four-way interaction was observed, $F(1, 187) = 14.07$, $p < .001$, $\eta^2_p = .07$, indicating that the effects of valence and validity on explicit and implicit evaluations was differentially moderated by the timing of validity information. To specify the particular nature of this interaction, analyses of explicit and implicit evaluations are reported separately for each of the two validity timing conditions.

3.3.2.3 Evaluations under short-delay validity timing

The condition involving a short delay before the presentation of validity information is conceptually identical to the design employed in Experiments 1 and 2, and analyses will proceed similarly. To test the effects of valence and validity feedback on explicit and implicit evaluations in the short-delay condition, standardized indices of explicit and implicit evaluations were submitted to a 2 (Valence: Positive vs. Negative) × 2 (Validity of Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) repeated measures ANOVA. A significant main effect was observed for valence, $F(1, 102) =$

10.51, $p = .002$, $\eta^2_p = .09$. In addition, significant two-way interactions were observed between valence and validity, $F(1, 102) = 561.63$, $p < .001$, $\eta^2_p = .85$, and between valence and evaluation type, $F(1, 102) = 4.14$, $p = .044$, $\eta^2_p = .04$. Finally, the three-way interaction between valence, validity, and evaluation type was significant, $F(1, 102) = 393.56$, $p < .001$, $\eta^2_p = .79$. No other effects were significant. As with Experiments 1 and 2, the qualification of the validity × valence interaction by evaluation type does not reflect the expected dissociation between explicit and implicit evaluations. Instead, the interaction reflected a slightly weaker effect size of the valence × validity cross-over interaction for implicit evaluations, as described below.

The effects of the valence and validity manipulations were assessed separately for both explicit and implicit evaluations using raw scores for all analyses. With respect to explicit evaluations, a significant main effect of valence was observed, $F(1, 102) = 17.98$, $p < .001$, $\eta^2_p = .15$, qualified by a significant two-way interaction between valence and validity, $F(1, 102) = 763.14$, $p < .001$, $\eta^2_p = .88$. As shown in Figure 3.5, validity information influenced explicit evaluations as expected, such that explicit evaluations reflected the valence of behavioral descriptions when it was true and the opposite valence when it was false. Paired-samples *t*-tests revealed that when the behavioral descriptions were true, explicit evaluations favored positively described targets over negatively described targets, $t(102) = 26.27$, $p < .001$; but when the behavioral descriptions were false, explicit evaluations favored negatively described targets over positively described targets, $t(102) = 20.92$, $p < .001$. Moreover, when the valence of behavioral descriptions was positive, explicit evaluations were more positive when the validity information was true rather than false, $t(102) = 21.52$, $p < .001$; but when the valence of behavioral descriptions was negative, explicit evaluations were more positive when the validity information was false rather than true, $t(102) = 25.08$, $p < .001$.

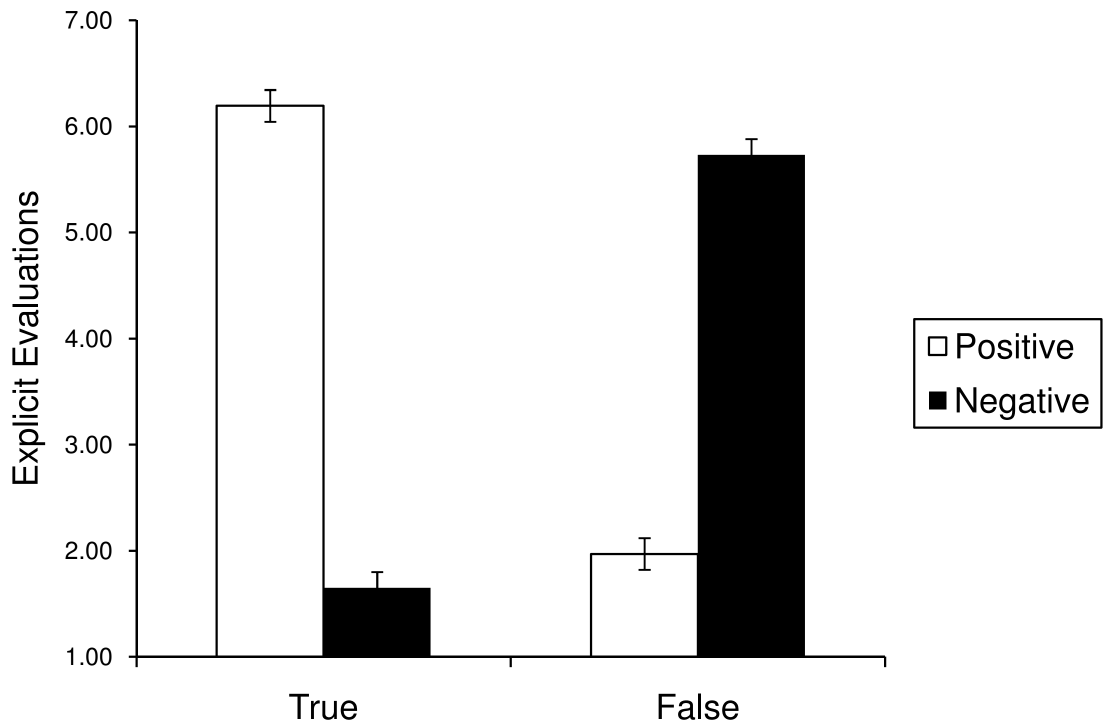

Figure 3.5. Explicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false) with short-delay validity feedback, Experiment 3.3. Error bars represent standard errors.

With respect to implicit evaluations, no main effects were significant but the two-way interaction between valence and validity was significant, $F(1, 102) = 32.37$, $p < .001$, $\eta^2_p = .24$. As shown in Figure 3.6, the pattern of the valence × validity interaction was identical to that obtained for explicit evaluations. Specifically, under quick validity feedback, implicit evaluations reflected the dominant valence when it was true and the opposite valence when it was false. Paired-samples *t*-tests revealed that when the behavioral descriptions were true, implicit evaluations favored positively described targets over negatively described targets, $t(102) = 4.62$, $p < .001$; but when the behavioral descriptions were false, implicit evaluations favored negatively described targets over positively described targets, $t(102) = 3.44$, $p = .001$. Moreover, when the valence of behavioral descriptions was positive, implicit evaluations were more positive when the validity information was true rather than false, $t(102) = 5.11$, $p < .001$; but when the valence of behavioral descriptions was negative, implicit evaluations were more positive when the validity information was false rather than true, $t(102) = 3.72$, $p < .001$.

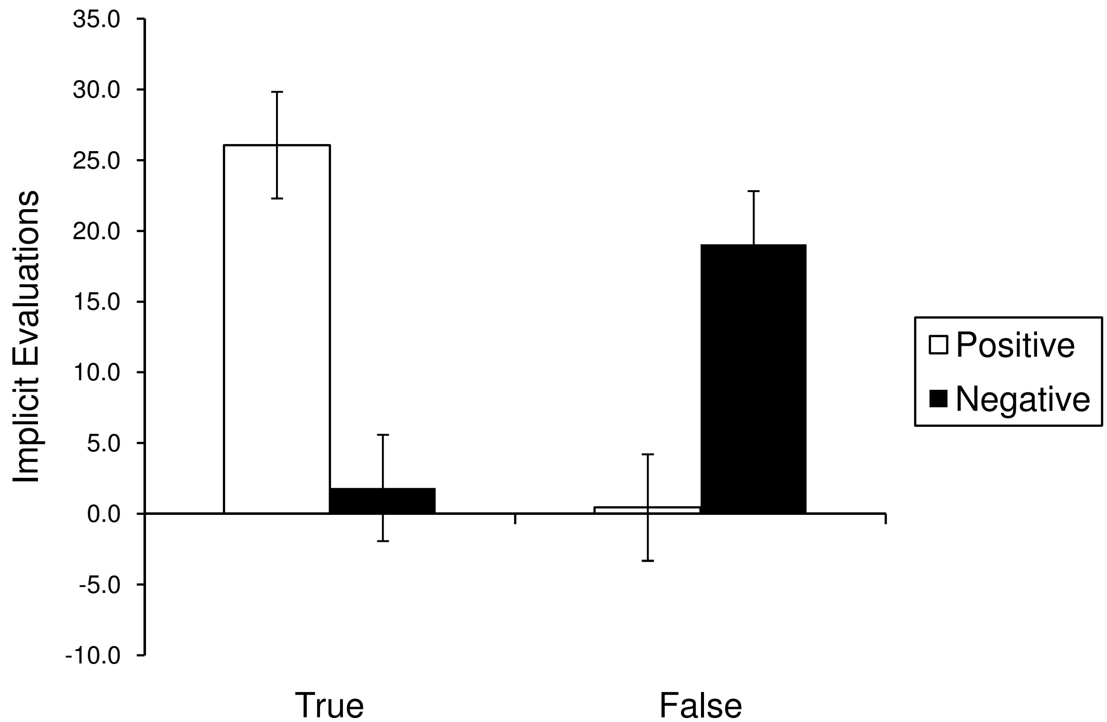

Figure 3.6. Implicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false) with short-delay validity feedback, Experiment 3.3. Error bars represent standard errors.

Taken together, these results replicate the findings of Experiments 1 and 2: When the delay between the presentation of valence and validity information was short, both explicit and implicit evaluations reflected the qualification of the behavioral descriptions by their perceived validity.

3.3.2.4 Evaluations under long-delay validity timing

To test for a potential dissociation between explicit and implicit evaluations in the long-delay condition, standardized indices of explicit and implicit evaluations were submitted to a 2 (Valence: Positive vs. Negative) × 2 (Validity of Valence: True vs. False) × 2 (Evaluation Type: Explicit vs. Implicit) repeated measures ANOVA. Significant main effects were observed for valence, $F(1, 85) = 44.15$, $p < .001$, $\eta^2_p = .34$, and validity, $F(1, 85) = 11.46$, $p = .001$, $\eta^2_p = .12$. In addition, significant two-way interactions were observed between valence and validity, $F(1, 85) = 161.51$, $p < .001$, $\eta^2_p = .66$, between valence and evaluation type, $F(1, 85) = 6.99$, $p = .010$, $\eta^2_p = .08$, and between validity and evaluation type, $F(1, 85) = 9.02$, $p = .004$, $\eta^2_p = .10$. Finally, the three-way interaction between valence, validity, and evaluation type was significant, $F(1, 85) = 172.05$, $p < .001$, $\eta^2_p = .67$. No other effects were significant. In this case, contrary to the results under quick validity timing, the qualification of the validity × valence effect by evaluation type does reflect a dissociation between explicit and implicit evaluations, as described below.

The effects of the valence and validity manipulations were assessed separately for both explicit and implicit evaluations using raw scores for all analyses. With respect to explicit evaluations, significant main effects were observed for valence, $F(1, 85) = 33.38$, $p < .001$, $\eta^2_p = .28$, and validity, $F(1, 85) = 19.60$, $p < .001$, $\eta^2_p = .19$, qualified by a significant two-way interaction, $F(1, 85) = 254.95$, $p < .001$, $\eta^2_p = .75$. As shown in Figure 3.7, validity information influenced explicit evaluations as expected, such that explicit evaluations reflected the valence of the behavioral descriptions when they were true and the opposite valence when they were false. Paired-samples *t*-tests revealed that when the behavioral descriptions were true, explicit evaluations favored positively described targets over negatively described targets, $t(85) = 16.72$, $p < .001$; but when the behavioral descriptions were false, explicit evaluations favored negatively described targets over positively described targets, $t(85) = 7.93$, $p < .001$. Moreover, when the valence of behavioral descriptions was positive, explicit evaluations were more positive when the validity information was true rather than false, $t(85) = 16.18$, $p < .001$; but when the valence of behavioral descriptions was negative, explicit evaluations were more positive when the validity information was false rather than true, $t(85) = 10.11$, $p < .001$.

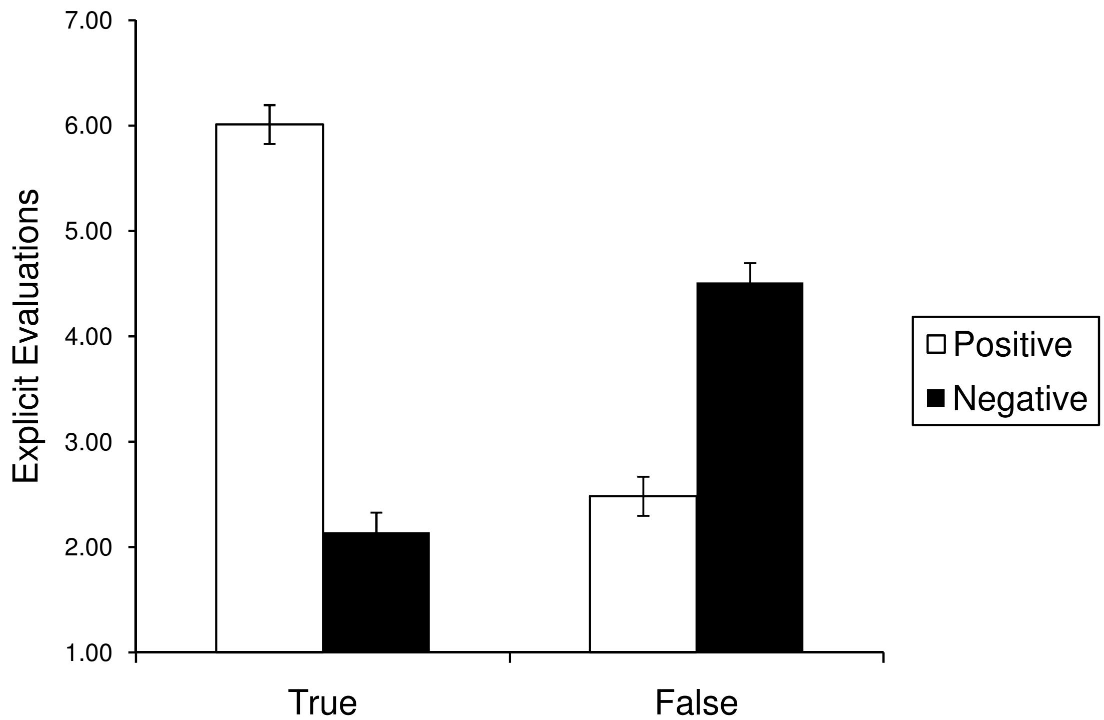

Figure 3.7. Explicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false) with long-delay validity feedback, Experiment 3.3. Error bars represent standard errors.

With respect to implicit evaluations, a significant main effect of valence was observed, $F(1, 85) = 9.75$, $p = .002$, $\eta^2_p = .10$, qualified by a significant two-way interaction between valence and validity, $F(1, 85) = 9.55$, $p = .003$, $\eta^2_p = .10$. No other effects were significant. As shown in Figure 3.8, the main effect of valence revealed that implicit evaluations of the positive targets were on average more positive than evaluations of the negative targets. This main effect of valence was qualified, however, by the validity of the behavioral descriptions, such that implicit evaluations reflected the valence of the behavioral descriptions when they turned out to be true, but this effect was only attenuated (rather than reversed) when the behavioral descriptions turned out to be false. Paired-samples *t*-tests revealed that when the behavioral descriptions were true, implicit evaluations favored positively described targets over negatively described targets, $t(85) = 38.88$, $p < .001$; but when the behavioral descriptions were false, implicit evaluations of negatively described targets were not significantly different from implicit evaluations of positively described targets, $t(85) = 0.31$, $p = .756$. Moreover, when the valence of behavioral descriptions was positive, implicit evaluations were more positive when the validity information was true rather than false, $t(85) = 2.37$, $p = .020$; but when the valence of behavioral descriptions was negative, implicit evaluations were more positive when the validity information was false rather than true, $t(85) = 2.25$, $p = .027$.

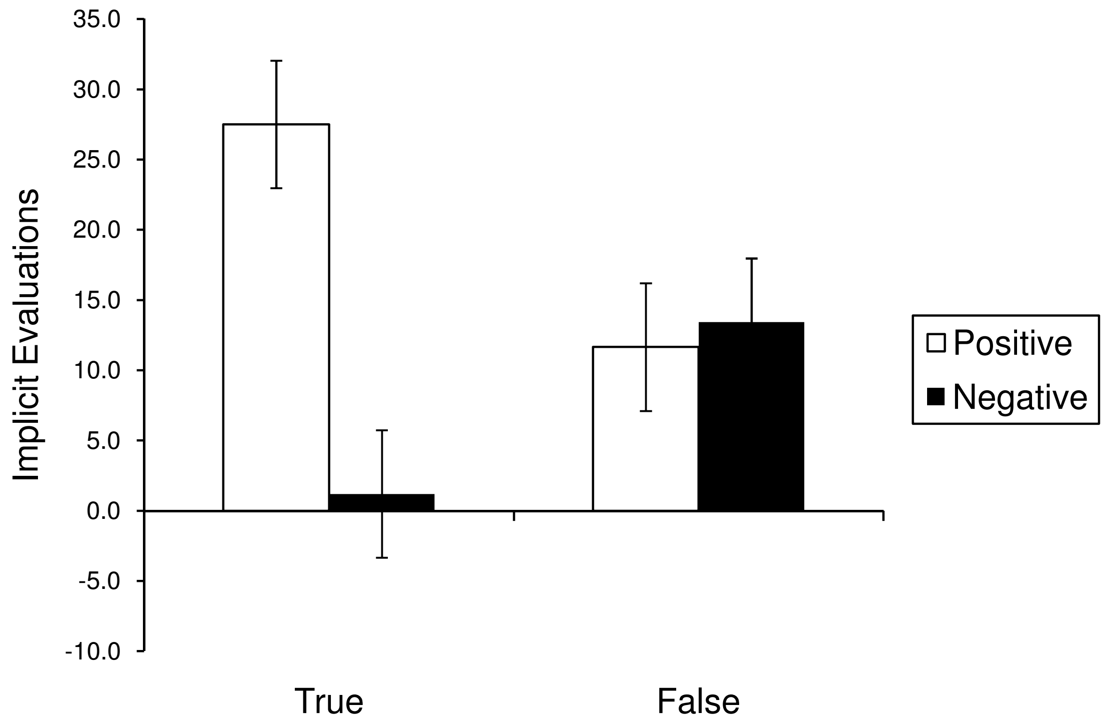

Figure 3.8. Implicit evaluations as a function of dominant valence (positive vs. negative) and validity of dominant valence (true vs. false) with long-delay validity feedback, Experiment 3.3. Error bars represent standard errors.

Thus, when the delay between the presentation of valence and validity information was relatively long, explicit and implicit evaluations became dissociated such that explicit evaluations reflected the full qualification of the behavioral descriptions by the validity information but effects on implicit evaluations were merely attenuated.

3.3.3 Discussion

The results of Experiment 3 support the hypothesis that evaluative dissociations may arise when the acquisition of validity information is delayed. When validity information was available during the learning of evaluative information, it qualified both explicit and implicit evaluations, replicating the results of Experiments 1 and 2. When the presentation of validity information was delayed, however, its impact was significantly reduced for implicit, but not explicit, evaluations. The current results thus imply a boundary condition on the emergence of evaluative dissociations, such that validity information may qualify both explicit and implicit evaluations when it is available during the acquisition of evaluative information; with the passage of time, however, changes in the perceived validity of previously acquired information may still qualify explicit evaluations, but will have a weaker effect on implicit evaluations. Evidence for asymmetric effects of validity information on explicit and implicit evaluations may therefore be explained as resulting from expression-related processes, rather than the simultaneous operation of two independent learning processes.

Although validity information had an asymmetric effect on implicit and explicit evaluations in the long delay condition, it is worth noting that it was still capable of partially qualifying implicit evaluations. Instead of reflecting the original valence of the behavioral descriptions, implicit evaluations did not differ between the two valence conditions when these descriptions were learned to be false. This attenuation deviates from Gregg et al.'s (2006) results, where validity information had no effect on implicit evaluations. Comparing the paradigms of the two sets of studies, there are at least two possible explanations for this difference. One explanation is that participants in the current experiments were forewarned that some information might turn out to be false, raising the possibility that our participants may have suspended belief in the observed contingencies until they knew their validity. To test this possibility, we conducted a replication of Experiment 3 in which participants were *not* informed, prior to the learning task, that some information might turn out to be false. The pattern of results was identical to that observed in Experiment 3, suggesting that the qualification of implicit evaluations observed in the current study is not due to the suspension of belief during learning. A second possible explanation is that Gregg et al.'s (2006) experiments involved a much longer delay between learning of evaluative information and subsequent acquisition of validity information. Whereas in our study, validity information was provided after all behavioral descriptions had been presented, participants in Gregg et al.'s studies completed measures of explicit and implicit evaluations before they were told that the initial behavioral information had been false. Thus, consistent with our emphasis on time as a critical factor, longer delays may allow consolidation of the initially formed associations. As a result, the impact of newly acquired validity information may decrease with increasing delays between the initial learning of evaluative information and the subsequent acquisition of validity information. Future research investigating the effects of continuously increasing intervals may help to clarify the role of time as a critical factor for the impact of validity information on implicit evaluations.

3.4 General Discussion

The present results provide converging evidence that during a single learning episode, in which the validity of the evaluation implied by an object-valence contingency can be quickly assessed, both explicit and implicit evaluations reflect a process of belief-formation rather than distinct effects of belief-based and contingency-based learning processes. In Experiments 1 and 2, the perceived validity of behavioral descriptions of social targets qualified both explicit and implicit evaluations of these targets. This result suggests that validity information is incorporated into the mental representation of evaluative objects at the time of learning. Expanding on these findings, Experiment 3 demonstrated that validity information can have asymmetric influences later at the time of expression, when evaluative responses are reconsidered in light of additional information. Manipulating the delay between the presentation of contingencies and the presentation of validity information revealed that the impact of validity information was reduced for implicit, but not explicit, evaluations when validity information became available after a substantial delay.

3.4.1 Implications for Dissociations between Explicit and Implicit Evaluations

Experimentally induced dissociations between explicit and implicit evaluations are often viewed as evidence for two independent learning mechanisms that may operate simultaneously on the basis of the same information (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004). The experiments reported here suggest that the best interpretation of evaluative dissociations depends on the specific conditions of the learning situation. Although there is compelling evidence that dissociations can arise during learning when the valence of an object is manipulated using multiple, distinct sources of information (e.g., Rydell et al., 2006), the present experiments found no evidence for the *simultaneous* operation of dual learning processes on the basis of the *same* information. In these situations, when the evaluations implied by observed contingencies are immediately qualified by validity information, both explicit and implicit evaluations seem to be driven by a single process of belief-formation.

Considering earlier evidence for evaluative dissociations arising during a single learning episode, it is worth noting that virtually all of this evidence can be straightforwardly interpreted as resulting from expression-related, rather than learning-related, processes. For example, using a learning procedure with a single, consciously available source of information, Ratliff and Nosek (2010) found that explicit evaluations showed the classic illusory correlation effect, whereas implicit evaluations reflected the actual contingencies of the observed information. Although they interpreted this finding as support for the independent operation of belief-based and contingency-based learning processes during a single learning episode, the dissociation can also be explained as a result of expression-related processes. In particular, illusory correlation effects may occur for explicit evaluations to the extent that infrequent information is more salient (Hamilton & Gifford, 1976), and salient information is given more weight in the process of generating an evaluative judgment. Importantly, such biases in the weighting of salient information may occur even if the relative strength of the underlying associations does not differ from the associations reflecting less salient information. From this perspective, the evaluative dissociation obtained by Ratliff and Nosek may not be due to the simultaneous operation of two learning mechanisms, but to processes operating during the generation of evaluative judgments.

3.4.2 Implications for Dual-Process Theories

The present findings seem, prima facie, more compatible with recent arguments for single-process theories of learning, according to which all learning is the product of a single process of belief-formation (e.g., Mitchell, De Houwer, & Lovibond, 2009). Evidence for dual-process learning in other situations notwithstanding, drawing such a conclusion in the present case seems premature for both empirical and epistemological reasons. First, it is always possible that in a single learning episode, the belief-based learning process is simply more powerful than the contingency-based learning process, thereby obscuring evidence for the operation of the latter process. Second, because theoretical entities cannot be observed directly, claims about their existence are not subject to direct empirical tests (Popper, 1934). Instead, existence claims have be evaluated indirectly by testing empirical predictions derived from assumptions about these entities. The underlying existence claims gain a measure of support when predictions are confirmed, but they will most likely be rejected when predictions repeatedly fail (Quine, 1969).

From this perspective, the current experiments can be understood as failing to confirm predictions derived from specific assumptions about contingency-based learning. The failure to confirm these predictions does not, however, conclusively demonstrate the non-existence of contingency-based, as opposed to belief-based, learning processes. Indeed, there is evidence for the operation of both contingency-based and belief-based learning processes in certain conditions, such that specific conditions may promote the operation of one process and inhibit the operation of the other (e.g., Rydell et al., 2006). The current findings do, however, challenge the idea that two learning processes operate *simultaneously* and lead to divergent explicit and implicit evaluations on the basis of the *same* information (e.g., Gawronski & Bodenhausen, 2006). In these situations, it appears that conscious beliefs exert a strong qualifying influence on the evaluations implied by observed contingencies at the time of learning. The current findings thus impose a constraint on the conditions under which dual learning processes may produce divergent outcomes, though they do not rule out their existence.

3.4.3 Implications for Mental Control

The findings of the present research also shed light on the question raised at the outset of this paper: Were Obama's supporters justified in worrying about an uncontrollable influence of negative images on voters? The answer seems to be: It depends. The experiments reported here suggest that beliefs about the invalidity of perceived object-valence contingencies (e.g., between *Obama* and *terrorism*) can qualify how that information is encoded and mentally represented, but only when the perceiver invalidates that information quickly. Otherwise, as the delay between the perception of a contingency and its invalidation increases, the mental association resulting from that contingency becomes more difficult to qualify (Gregg et al., 2006; Petty et al., 2006). Thus, if a potential voter observed the illustration of Obama as a terrorist, she might be able to prevent the depicted link from being stored in memory by immediately rejecting it as false; failing to do so, however, might lead to the formation of a mental association that directly reflects it. This outcome would indeed be worrisome for Obama's supporters, as implicit evaluations have been shown to predict significant behavioral outcomes, including choice decisions in the political domain (e.g., Galdi et al., 2008; Payne et al., 2010). Nevertheless, there seems to be some room for control over the evaluations we form in typical learning situations, such that contingency-based learning may not be powerful enough to create mental associations that contradict our beliefs about what we observe. If we are puppets on a string to those who would seek to influence us, we at least have a brief opportunity to pull back.

3.5 References

Banaji, M. R. (2008). The science of satire: Cognition studies clash with 'New Yorker' rationale. *Chronicle of Higher Education, 54,* B13 (August 1, 2008).

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

Carraro, L., Gawronski, B., & Castelli, L. (2010). Losing on all fronts: The effects of negative versus positive person-based campaigns on implicit and explicit evaluations of political candidates. *British Journal of Social Psychology, 49,* 453-470.

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure*. Journal of Experimental Social Psychology, 45,* 101-114.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of Automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385-405.

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

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

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

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

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731.

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

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34,* 1276-1289.

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

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

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12,* 392-407.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16*, 335-390.

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

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

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology, 46,* 367-374.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST Model. *Journal of Personality and Social Psychology, 90,* 21-41.

Popper, K. R. (1934). *Logic of scientific discovery*. New York: Basic Books.

Quine, W. V. O. (1969). Ontological relativity. In *Ontological relativity and other essays* (pp. 26-68). New York: ColumbiaUniversity Press.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510-532.

Ratliff, K. A., & Nosek, B. A. (2010). Creating distinct implicit and explicit attitudes with an illusory correlation paradigm. *Journal of Experimental Social Psychology, 46,* 721-728.

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*, 954-958.

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

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

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

4 General Discussion

In this chapter, I will explore a new approach to carving cognition in light of the challenges the present research poses for existing approaches. I will first work up to a specification of this new version of the cognitive covariation thesis by considering a distinction between two kinds of mental representation. Second, I will discuss the implications of this new version of the thesis for the problem of measurement, both with regard to the reinterpretation of existing measurement procedures and the design of new procedures. Finally, I will describe directions for future research implied by the new versions of the cognitive and psychometric covariation theses developed here.

4.1 A new approach to carving cognition

The findings from the present research underscore the empirical challenges facing the contextualization and validation hypotheses. The self-construal research reported in §2 provides further evidence for top-down influences on implicit cognition; this finding is consistent with the validation hypothesis, but contradicts the assumption of the contextualization hypothesis that the direction of influence between processes is essentially bottom-up. The evaluative learning research reported in §3 suggests that implicit cognition is, under certain conditions, sensitive to perceived validity; this finding is consistent with the contextualization hypothesis, but contradicts the assumption of the validation hypothesis that only higher-level cognition is sensitive to perceived validity. Together with the difficulties addressed in the introduction, these findings call into question the accuracy of the underlying monitoring-and-correction processing schema for modeling human cognition: Perhaps, despite its folk psychological appeal, the characterization of lower-level cognition as requiring higher-level correction to be made situationally appropriate is simply incorrect.

In this section, I will sketch an alternative approach to carving cognition that does not rely upon the monitoring-and-correction schema. The basic insight can be found with a closer examination of the studies of evaluative learning presented in §3, particularly Experiment 3. The original conclusion drawn from this experiment was that temporal delay weakened the top-down influence of negation on implicit evaluations and thereby produced a dissociation between explicit and implicit evaluations. It is not immediately clear, however, why time between the encoding of valence information and the encoding of validity information should asymmetrically influence these evaluations—in other words, why implicit evaluations (compared to explicit evaluations) should show such a strong primacy effect (Gregg, Seibt, & Banaji, 2006). The validation hypothesis, as described in the APE model (Gawronski & Bodenhausen, 2006, in press) and the RIM (Strack & Deutsch, 2004), provides no inherent reason why implicit evaluations should show a stronger primacy effect than explicit evaluations, especially given that implicit evaluations were sensitive to validity information when it was available without delay (see Experiments 3.1 and 3.2). The contextualization hypothesis might provide a basis for this prediction based on the assumptions that, once learned, the simple object-valence associations underlying implicit evaluations are less sensitive to contextual information in general, and that when validity information was presented after a delay it was somehow more "contextual" than when it was presented immediately during the learning task. Yet the assumption that lower-level, implicit processing is generally less context-sensitive than higher-level, explicit processing has not been supported by empirical evidence (see §1.2.1.2); for example, the experiments on self-construal in §2 demonstrated that contextual changes in the accessibility of information (Experiment 2.1) or in motivational orientation (Experiment 2.2) effectively influenced the implicit self-concept. It is therefore difficult to argue that the delayed presentation of validity information had an asymmetric effect solely because it created a stronger distinction between object and context.

Neither of the major approaches to dual-process social-cognition, then, provides a satisfactory account of why the temporal delay in Experiment 3.3 uniquely weakened the influence of validity on implicit evaluations. A promising alternative approach, however, is based on the insight that in this experiment, temporal delay was confounded with *quantification*, by which I mean the notion of quantifying across a number of tokens by grouping them into a single type. This is in fact an accurate description of the delayed validity manipulation in Experiment 3.3, in which participants were required to quantify across all of the object-valence associations they had previously observed for each target by labeling the whole group as true or false. This manipulation required participants to create a new type-level representation for each of the four targets, comprising each of the (token-level) object-valence associations observed in the learning task. Crucially, it was the evaluation of this quantified, *type-level* representation that participants in this condition were asked to negate, rather than each individual (token-level) object-valence association. Reversing learned evaluations after a delay, therefore, amounted to *quantified negation*, and this may be the reason that implicit evaluations were insensitive to perceived validity after the delay: The key distinction between higher-level and lower-level cognition may be the capacity for quantification, such that lower-level cognition represents unique tokens, whereas higher-level cognition is capable of grouping these tokens into abstract types and manipulating these abstractions via *quantified* operations.

4.1.1 Grounding the cognitive covariation thesis in mental representation

In order to state this quantification hypothesis more clearly as a specification of the cognitive covariation thesis (i.e., in terms of operating conditions and operating principles), it will help to clarify first why the capacity for quantification might be expected to distinguish between kinds of cognition. In the earlier discussion of the unity problem (§1.3), the monitoring-and-correction processing schema was criticized as lacking a reasonable explanation for distinguishing between kinds of cognition in terms of sensitivity to the situational context. The weakness of the various explanations, however, was not a result of the appeal to context-sensitivity itself but to the manner in which that distinction was applied. In particular, both the contextualization and validation hypotheses view higher-level processing as *more* sensitive to situational context (either in general or with respect to situational consistency) than lower-level processing, the basic idea being that higher-level processing "corrects" lower-level processing by taking aspects of the situational context into account. The architectural explanation attempts to derive this distinction from differences in mental representation, such that similarity-based processing in connectionist networks is corrected by rule-based processing in classical architectures. When mental representation is assumed to be exclusively compositional, however, the classical/connectionist distinction fails to justify this argument: Compositionality requires that representation in both classical and connectionist architectures be localist (i.e., represent a discrete object or concept), with the consequence that neither architecture is *inherently* more context-sensitive. Likewise, the phylogenetic explanation for viewing higher-level processing as more situationally sensitive is unconvincing. Insofar as quick responses need to be evolutionarily adaptive, early cognition must be more, not less, sensitive to the situational context (Schwarz, 2007); carving percepts into discrete stimuli and sorting these into abstract categories would seem to be luxuries afforded by time and cognitive capacity.

As anticipated earlier (see §1.2.3), the basic problem with the monitoring-and-correction processing schema may be that it rests on the assumption that mental representation is exclusively compositional—that is, that representations in one kind of process are completely reducible to those in another kind of process. Indeed, if the contrast between classical and connectionist architectures has any value for dual-process theories, it is that it makes possible a distinction between localist, compositional representation and distributed, non-compositional representation. This distinction between kinds of mental representation, moreover, is capable of grounding a distinction between kinds of cognitive process in terms of context-sensitivity, though in a way fundamentally different from that envisioned by the monitoring-and-correction processing schema. Specifically, distributed representation is not just context-sensitive but context-*dependent*: Not being composed from discrete units of meaning, distributed representations lack a clear distinction between object and context; instead, such representations are *holistic* in that the object of representation is a unique "object-in-context" (Chrisley, 1998; Ferguson & Bargh, 2007; Smolensky, 1988; cf. Fodor, 1987; Fodor & Lepore, 1992). The major consequence for present purposes is that, because holistic representations lack discrete units of meaning, they are fundamentally non-conceptual and non-linguistic—and therefore cannot be reduced to localist representations in a classical architecture. Indeed, the contrast between these two kinds of mental representation has long been recognized as central to cognitivist psychology. James (1890) devoted an entire chapter of his *Principles* to describing the holistic "stream of thought," and he placed heavy emphasis on the notion that cognition is not altogether discrete and linguistic:

> *Language works against our perception of the truth. We name our thoughts simply, each after its thing, as if each knew its own thing and nothing else. What each really knows is clearly the thing it is named for, with dimly perhaps a thousand other things. It ought to be named after all of them, but it never is. Some of them are always things known a moment ago more clearly; others are things to be known more clearly a moment hence. (p. 241)*

I will refer to this distinction between kinds of mental representation, and the kinds of cognitive process that operate upon them, as *holistic* and *symbolic*.[12] Holistic representation is distributed, which entails that it is non-compositional, non-linguistic, and context-dependent. Holistic processing is assumed to be realized by distributed connectionist architectures that operate according to principles of similarity and contiguity across patterns of activation (Smolensky, 1988). In contrast, symbolic representation is localist, which entails that it is compositional, linguistic, and context-independent. Symbolic processing is assumed to be realized by classical architectures that are essentially rule-based Turing machines (i.e., manipulators of abstract symbols), per the Computational Theory of Mind (Horst, 2009; Newell, 1980).

By virtue of the underlying distinction between kinds of mental representation, the relation between these two kinds of cognition involves quantification: Unique, holistic tokens at the lower level may be abstracted into symbolic types at the higher level.[13]

---

[12] A brief note on nomenclature: Nisbett, Peng, Choi, and Norenzayan (2001) distinguish between *analytic* and *holistic* cognition, whereas Smolensky (1988) distinguishes between *symbolic* and *subsymbolic* cognition. With regard to the former case, I have chosen to use the term *symbolic* rather than *analytic* because it gives a clearer sense of the representational assumptions underlying this kind of cognition. In the latter case, I have chosen to use the term *holistic* rather than *subsymbolic* because is offers a direct, rather than relative, characterization of the representational assumptions underlying this kind of cognition.

[13] With its emphasis on the relation between abstract type-level representations and unique token-level representations, the quantification hypothesis might be seen as similar to existing social-cognitive theories that take the "level of representation" of an object into account (e.g., Lord & Lepper, 1999; Trope & Liberman, 2003; Vallacher & Wegner, 1987). This similarity is only superficial, however, given that these theories assume that all representations are localist, though they can differ in degrees of abstraction. The quantification hypothesis, in contrast, assumes a qualitative distinction between kinds of representation (i.e., holistic vs. symbolic), meaning that representations at different levels are not reducible. Thus, for

Thus, quantification across context-dependent tokens creates context-independent types that can be processed in a classical architecture. In formal logic, the relevant notion of quantification is specifically *universal quantification*, denoted $\forall$. The capacity assumed to be the hallmark of higher-level, symbolic cognition is therefore any quantified predication of the form $\forall(x)Fx$, which is read "for all tokens of type x, x is F." In this sense, types are mental categories (i.e., concepts) that subsume similar tokens. For example, in the statement "Dogs are red," it is obvious that *dogs* is a category. Yet even in predications of singular concepts, such as in the statement "My dog is red," *my dog* is equally categorical, for the reason that the predication subsumes all *representational instances* or *tokens* of my dog in memory or perception; essentially, *my dog* is a type rather than a token in the current perspective because it is *repeatable*. Quantified predication is therefore involved in any attribution of a property to a linguistic concept (i.e., a noun), which is obviously a frequent occurrence in human cognition. Conversely, holistic processing in connectionist architectures is incapable of quantification, because distributed representations cannot be treated as discrete, repeatable types. Once again, James (1890) gave lucid expression to this point:

> *Every thought we have of a given fact is, strictly speaking, unique, and only bears a resemblance of kind with our other thoughts of the same fact. When the identical fact recurs, we must think of it in a fresh manner, see it under a somewhat different angle, apprehend it in different relations from those in which it last appeared. And the thought by which we cognize it is the thought of it-in-those-relations, a thought suffused with the consciousness of all that dim context. (p. 233)*

Although quantification is a powerful cognitive capacity, its use involves a tradeoff between representational flexibility and sensitivity to the immediate situation. Symbolic,

---

example, token-level representations in Trope and Liberman's (2003) construal level theory would nevertheless be viewed as type-level, abstract representations from the perspective of the quantification hypothesis. Most importantly, *tokens* in the current perspective are assumed to be represented holistically.

type-based processing allows for the consideration of counterfactuals, such that alternatives to the immediate situation can be conceived via context-independent representation (Roese & Olson, 1995); yet the symbolic processing that makes counterfactual cognition possible also requires that the holistic meaning of the immediate situation to the organism is lost. Essentially, the tradeoff is that symbolic processing facilitates long-range planning, but requires a compositional reconstruction of the immediate situation in terms of discrete, quantifiable types (i.e., linguistic concepts) that can be mentally manipulated to simulate various contingencies; holistic processing facilitates immediate, context-sensitive responses, but being context-dependent cannot perform such counterfactual manipulations.

In light of this tradeoff, it is important to note that neither kind of cognition is inherently normative from the perspective of the quantification hypothesis. The monitoring-and-correction processing schema locates the distinction between kinds of cognition in their relative complexity, such that higher-level processing is assumed to be more sensitive to situational context (either in general or with respect to situational consistency). The implication of this schema is that higher-level processing will always be more "situationally appropriate," because it is able to take more situational information into account. In contrast, the distinction between holistic and symbolic cognition avoids this normative implication, because it does not locate the distinction in the relative amount of information considered but in the *kinds* of information considered, which serve complementary functions for an organism. Thus, holistic representations, being context-dependent, make possible behavioral responses that are acutely sensitive to immediate contingencies. Symbolic representations, being context-independent, make possible behavioral responses that can take counterfactual contingencies into account. Neither kind of cognition, however, is inherently more normative than the other, as each serves distinct goals that are potentially adaptive for an organism (cf. Strack & Deutsch, 2004). Indeed, this perspective offers a more nuanced view of normativity in psychology by recognizing that the "situational appropriateness" of a behavioral response can be understood in terms of both immediate and long-term contingencies, and that both can serve adaptive functions for an organism. Which of these two responses might be more "correct" when they conflict cannot be decided a priori, based on the principles by which holistic and symbolic cognition operate, but is rather a question that depends on external assumptions about what constitutes normative behavior (Rey, 2007). In this sense, the quantification hypothesis altogether dispenses with the monitoring-and-correction processing schema for modeling human cognition, which has informed the vast majority of dual-process research in social-cognitive psychology.

We are now in a position to return to the question raised at the beginning of this section: Why might the capacity for quantification be expected to distinguish between kinds of cognition? In contrast to the contextualization and validation hypotheses, the quantification hypothesis is grounded in a distinction between kinds of mental representation that provides the foundation for reasonable responses to the unity problem. First, the architectural and neurophysiological explanations become defensible with the distinction between genuinely different kinds of mental representation (cf. §§1.3.1-1.3.2). As discussed above, the relation between holistic and symbolic representations can be described as a process of quantification. Thus, if the human mind is characterized by *these* two kinds of representation in particular, it makes sense that the capacity for quantification would be the key distinction between kinds of cognition. Second, the quantification hypothesis is more amenable to a phylogenetic explanation. As discussed earlier (§1.3.3), one problem with the monitoring-and-correction processing schema is that it makes little sense that the behavior of lower animals, which presumably lack higher-level cognitive capacities, would be evolutionarily adaptive if it were based on normatively deficient cognition. In particular, the idea that early, resource-efficient cognitive processing is insensitive to immediate contingencies, including situational consistency, is difficult to square with the fact that lower animals are capable of manifestly intelligent behavior. The quantification hypothesis suggests, instead, that it is not the amount of information guiding behavior that distinguishes between different kinds of cognition, but the kind of information: Thus, behavior in lower animals should be guided more or less exclusively by similarity-based processing across holistic representations, a contention supported by comparative research on category learning and use (e.g., Smith, Minda, & Washburn, 2004; Smith et al., 2011). Such context-dependent cognition would be eminently adaptive in the short term, but would lack the capacity for counterfactual representation that makes long-range planning possible. It may not be coincidental that this capacity for "decoupling" cognition from the immediate situation is frequently identified as the hallmark of human sentience (e.g., Cosmides & Tooby, 2000; Dienes & Perner, 1999; Stanovich, 2004).

These considerations offer a final push toward a precise specification of the quantification hypothesis. In particular, they suggest that the conditions under which holistic and symbolic processes operate can be characterized with respect to their complementary functions: Holistic processing, though tied to the immediate situation, facilitates context-sensitive responding much faster than would be possible if the situation were reconstructed compositionally, in terms of discrete concepts; symbolic processing, on the other hand, does depend on such an effortful reconstruction (via quantification), but the decomposition of a holistic representation into discrete, context-independent concepts makes possible counterfactual thinking. Based on this characterization of holistic and symbolic cognition, the key distinction between their respective operating conditions would appear to be the *sensory-dependence* of cognitive processing. Thus, holistic cognition should be expected to operate under conditions of high sensory-dependence, whereas symbolic cognition should be expected to operate under conditions of low sensory-dependence—in other words, when a person has time to "think about" his or her response. Indeed, this colloquial expression is intended quite literally here, as the proposed relation between holistic and symbolic cognition is essentially metacognitive: Holistic cognition can be considered first-order, experiential cognition, whereas symbolic cognition involves thinking about these first-order processes (Dienes & Perner, 1999; Rosenthal, 2002). It is this capacity for discretizing holistic representations—essentially, abstracting repeatable types from unique tokens—that is the hallmark of symbolic cognition.

The relation between high and low sensory-dependence and the more traditional distinction between automaticity and control (e.g., Bargh, 1994) is likely not isomorphic, and this is an important area for future research (see §4.2.1 for further discussion). For the present, I will simply emphasize that sensory-independence should require time and capacity to decouple cognition from perception (i.e., to make the leap from experiential to metacognitive processing). A rough assumption, then, is that (all else being equal) early responses will be highly sensory-dependent and later responses should be less so. Furthermore, in specifying the operating conditions of the quantification hypothesis, I will use the terms *online* and *offline* to refer to conditions of high and low sensory-dependence, respectively (for similar uses of these terms, see Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005, and Wilson, 2002). One reason for doing so is that these terms succinctly capture the meaning of the proposed distinction between operating conditions, such that cognitive processing under high sensory-dependence is directly connected to experience, whereas processing under low sensory-dependence is to some degree decoupled from experience as a result of quantification. Second, the online/offline nomenclature avoids the normative connotations inherent in the distinction between automaticity and control, and thus carries no implications about which operating conditions are more likely to produce more adaptive responses.

4.1.2 The quantification hypothesis

The quantification hypothesis is presented in Table 4.1 as a proposed specification of the cognitive covariation thesis.

The Quantification Hypothesis	
Operating Conditions	
Online	Offline
High sensory-dependence (“Experiential”)	Low sensory-dependence (“Metacognitive”)
Operating Principles	
Holistic	Symbolic
Distributed/Connectionist*	Localist/Classical*
Situation-dependent	Counterfactual
Continuous	Discrete

Table 4.1. The specification of the cognitive covariation thesis per the quantification hypothesis.
*Denotes the key distinction between cognitive operating principles.

The basic claim of the quantification hypothesis is that online cognition is characterized by the principles of holistic processing, whereas offline cognition is characterized by the principles of symbolic processing. The hypothesis assumes that these two kinds of cognition are related via a process of quantification, in which holistic representations are decoupled from immediate experience to produce abstract, context-independent concepts that can be manipulated via quantified operations (e.g., linguistic syntax, formal logic, etc.). This perspective suggests that the distinctions drawn by the monitoring-and-correction approach to dual-process cognition (i.e., in terms of automaticity/control and situational sensitivity; Table 1.1) were on the right track, but the quantification hypothesis turns this approach on its head by suggesting that early processing should be *more* sensitive to situational context rather than less so (cf. §§1.2.1.1, 1.2.2.1). According to this new approach, generality is an effortful cognitive achievement, not the starting point of cognitive processing (Ferguson & Bargh, 2007).

4.1.2.1 Sources of influence on holistic and symbolic cognition

The quantification hypothesis implies that, due to the nature of distributed vs. localist representation, the early meaning of a stimulus to an individual should be context-dependent and later meaning should be less so. This suggests how the current scattershot approach to investigating the context-sensitivity of implicit cognition—which essentially involves testing whether a given contextual factor (e.g., visual background, active goals, physical environment, etc.) influences implicit processing—might be brought into focus: by recognizing that it is not *what* is manipulated, but *how*, that is responsible for asymmetric effects on higher-level and lower-level processing. In particular, the symbolic meaning of a stimulus can be directly assigned, independent of context, simply by relabeling a symbol to mean something else; in other words, symbolic cognition is capable of *predication*, as described in §4.1.1. In contrast, holistic meaning is determined by the situational context and cannot simply be changed through reassignment, since there is no discrete symbol upon which the (quantified) reassignment can operate. Consequently, the quantification hypothesis predicts that attempts to change the meaning of a stimulus through direct reassignment should affect symbolic (explicit) processing but not holistic (implicit) processing; conversely, attempts to change meaning by manipulating the context of a stimulus should affect holistic (implicit) processing but not symbolic (explicit) processing.

These predictions are borne out by existing research on the context-sensitivity of implicit processing (see §1.2.1.2), but go further by providing a general rationale that ties these varied findings together: Manipulations of *context* directly influence early, context-dependent meaning, whereas *reassignments* of meaning directly influence late, context-independent meaning. An experiment by Foroni and Mayr (2005) provides an excellent illustration of this point. When participants were simply told to think of insects as positive and flowers as negative, implicit evaluations of insects and flowers were unaffected, with insects being more strongly associated with negativity than flowers on an IAT (see also Gregg et al., 2006). This is a standard finding and is assumed to reflect the evaluation associated with each stereotype (i.e., insects are bad, flowers are good; Greenwald, McGhee, & Schwartz, 1998). In another condition, however, participants were asked to imagine that they were searching for food in a post-apocalyptic world in which flowers are highly radioactive and insects provide the only safe source of sustenance. This manipulation, which changed the meaning of the targets via the context in which they were perceived, successfully reversed implicit evaluations, such that flowers were more strongly associated with negativity than insects. In general, then, lower-level processing might be usefully understood as reflecting the contextual meaning of a stimulus, whereas higher-level processing reflects its decontextualized (assigned) meaning.

From this perspective, the current, somewhat atheoretical approach to investigating the context-sensitivity of implicit cognition can be viewed more coherently as a program seeking to identify the aspects of situational context that are *relevant* to the determination of holistic meaning—and it seems there are indeed many (see §1.2.1.2). This perspective, moreover, conceptually integrates the explosion of findings within the "situated" and "embodied" cognition literatures, in which a supposedly irrelevant aspect of the situational context systematically influences behavior (Niedenthal et al., 2005; Smith & Semin, 2004). For example, Williams and Bargh (2008) found that holding a warm (vs. cold) mug of coffee increased perceptions of interpersonal warmth, despite the fact that participants reported no awareness of this influence. Lee and Schwartz (2010) found that asking participants to wash their hands (ostensibly as part of an unrelated product testing study) reduced feelings of regret after making a choice between two equally attractive alternatives. Although such findings are generally regarded as novel, they appear much less surprising when it is simply assumed that holistic meaning (which presumably mediates the behavioral outcomes in such experiments) is dependent on the situational context. Put the other way around, these findings are counterintuitive only if it is assumed that the meaning of a stimulus is always context-independent and that the only contextual factors that might influence meaning are those that have been explicitly identified. Although such computer-like cognition might be viewed as the ideal (at least within analytic Western cultures; Nisbett, Peng, Choi, & Norenzayan, 2001), research in social psychology has long made clear that cognition is heavily influenced by context, regardless of whether it is explicitly identified or not (Markus, 2005; Nisbett & Wilson, 1977). For example, Schwarz and Clore (1983) famously found that reports of life-satisfaction were systematically influenced by the weather at the time of the interview. Likewise, the extensive literature on affective forecasting errors reveals that people have a difficult time controlling for the influence of the immediate context, and taking counterfactual contexts into account, when predicting their behavior in future situations (Wilson & Gilbert, 2003).

The current fascination with "situated" effects in social psychology is due largely to an experiment by Bargh, Chen, and Burrows (1996), the history of which helps to illuminate the broader point being made here. In this experiment, participants surreptitiously primed with words related to the "elderly" stereotype were found to walk more slowly after leaving the laboratory. Bargh et al. interpreted this finding as evidence for the activation of the "elderly" concept (i.e., stereotype) in memory, which had a direct influence on behavior. In later work, however, Cesario, Plaks, and Higgins (2006) found that this effect was moderated by individual differences in participants' associations with the elderly, such that participants with positive associations walked more slowly after being primed whereas participants with negative associations walked more quickly. Cesario et al. concluded that these behavioral differences could be understood as a result of participants' preparation to interact with the primed group; in other words, participants with more positive associations behaved more sympathetically, presumably to smooth social interaction (see also Sinclair, Lowery, Hardin, & Colangelo, 2005). Contrary to Bargh et al.'s (1996) interpretation, this conclusion suggests that "priming" via the manipulation of supposedly irrelevant aspects of the situational context does not activate a context-independent concept, or *stereotype*, but rather activates the *meaning of the prime within that specific situation*—that is, a holistic, rather than symbolic, meaning (for similar accounts, see Perugini & Prestwich, 2007; Wheeler, DeMarree, & Petty, 2007).

4.1.2.2 Directions of influence

So far, I have focused on the relation between holistic and symbolic cognition in terms of a *bottom-up* process of quantification. Thus, the evaluative dissociation observed in Experiment 3.3, which resulted from the introduction of a delay between the learning of object-valence associations and their validity, may be more fruitfully interpreted as the result of the unique influence of quantified negation on symbolic representations, which have been abstracted from lower-level, holistic representations. Yet it is clear that top-down influences can also occur, such that changes in directly measured cognition appear to mediate changes in indirectly measured cognition (Gawronski & Bodenhausen, 2006). From the perspective of the quantification hypothesis, Experiments 3.1 and 3.2, which involved a short delay between the learning of object-valence associations and their validity, would not in fact be interpreted as top-down effects, since the hypothesis allows for the direct influence of perceived validity at both levels; the critical difference is whether or not the application of the truth-value is quantified, and in these two experiments this was not the case.

Experiment 2.2 does, however, provide a good example of how a top-down influence can be understood from the perspective of the quantification hypothesis. In that experiment, participants were motivated to assert an abstract belief about themselves (e.g., "I am extraverted"). Essentially, this belief can be understood as a quantified affirmation, such that it asserts the validity of any and all token-level self-knowledge that confirms the type-level belief. In order to substantiate this belief, participants were presumably motivated to engage in a search through token-level self-knowledge in memory biased toward confirmation (Kunda, 1990). Thus, from the perspective of the quantification hypothesis, top-down influences can be broadly conceptualized as "deductive" in the sense that higher-level assertions serve as premises that lead to the activation of specific lower-level knowledge. Conversely, the bottom-up route of self-construal observed in Experiment 2.1 can be understood as an "inductive" relation between type-level and token-level processes. In this case, an abstract, type-level belief (e.g., "I am extraverted") is induced from the prior activation of multiple tokens of self-knowledge (e.g., specific memories of one's own extraverted behaviors).

4.1.3 Summary

The quantification hypothesis is proposed as a response to the conceptual and empirical challenges facing the contextualization and validation hypotheses. The quantification hypothesis is grounded in a distinction between genuinely different (irreducible) kinds of mental representation, and the properties of each kind of representation that are assumed to inform the operating principles of cognition: Specifically, distributed representations underlie holistic processing whereas localist representations underlie symbolic processing. The process of translating between these two kinds of representation—referred to as quantification—is assumed to characterize dual-process interactions, both in a bottom-up, inductive direction and a top-down, deductive direction. Moreover, the hypothesis suggests that holistic processing should be directly influenced by changes in situational context that affect the meaning of a stimulus, whereas symbolic processing should be directly influenced by reassignments of meaning to a stimulus. The operating conditions of the quantification hypothesis are derived from the complementary functions of each kind of cognition, which serve to generate responses based on either immediate situational contingencies or counterfactual contingencies. Thus, online operating conditions are highly sensory-dependent ("experiential"), whereas offline operating conditions are less sensory-dependent ("metacognitive").

4.2 The measurement problem: Leveraging the cognitive covariation thesis

Dual-process theories of cognition do not just provide a basis for explanations of human behavior—they also provide the basis for the design and interpretation of the measurement procedures necessary for testing such explanations. As noted in §1.1, this dialectical relation between observed behavior and explanatory theory can be approached from either direction, and in the present discussion I have focused initially on the theoretical assumptions used to characterize different kinds of cognition. Beyond offering a new cognitive framework for explaining behavior, however, this hypothesis also sheds light on the enduring problem of measurement: That is, what are our measurement procedures measuring? In this way, the specification of the cognitive covariation thesis can be leveraged to refine the specification of the psychometric covariation thesis (Figure 1.2). In this section, I will first address the prospects for reinterpreting existing measurement procedures in light of the quantification hypothesis and then discuss implications for the development of new measurement procedures. Finally, I will explore the possibility that the distinctions between the operating principles and operating conditions of cognition are not qualitative but *graded*.

4.2.1 Reinterpreting existing measurement procedures

In the received view of dual-process cognition, direct and indirect measurement procedures are classified in terms of the operating conditions of automaticity and control, and existing measurement procedures have generally been designed with this distinction in mind. This complicates the application of the quantification hypothesis to the reinterpretation of existing measurement procedures, since this hypothesis draws an alternative distinction between operating conditions in terms of online and offline processing. Thus, the first obstacle to reinterpreting these measurement procedures is to determine the degree to which the respective operating conditions overlap: In particular, how well does the automatic/controlled distinction map on to the online/offline distinction? This is a complicated question in itself, given that many functional properties of automaticity have been proposed without a clear consensus in the social-cognitive literature (Bargh, 1992, 1994; Moors & De Houwer, 2006). For example, some indirect measurement procedures are designed to emphasize distinctions in conscious awareness (e.g., subliminal priming), whereas others are designed to emphasize distinctions in resource-efficiency (e.g., through the use of a distracter task). Moreover, as mathematical modeling has revealed, it is unclear how distinct the proposed features of automaticity and control actually are and how well measurement procedures can selectively establish them (Sherman, Klauer, & Allen, 2010).

Nevertheless, there is at least a family resemblance between the automatic/controlled and online/offline distinctions, which might best be captured by the overlap between the features of *resource-efficiency* and *sensory-dependence*. The link between these two features is that the cognitive resources upon which resource-heavy processes depend may be precisely those that are required to take online, sensory-dependent cognition offline—that is, via the (effortful) quantification of holistic representations. From this perspective, existing measurement procedures that emphasize resource-efficiency (e.g., through the use of a distracter task) might be most easily reinterpreted with the quantification hypothesis, insofar as they discriminate between the conditions upon which the capacity for quantification depends. If this is the case, then the quantification hypothesis implies that such measurement procedures reflect the operation of holistic cognition, rather than "implicit" cognition as characterized by the contextualization and validation hypotheses (Table 1.1 and Table 1.2). This would mean that such measurement procedures should be influenced by situational contingencies in general, including perceived validity, but should be specifically insensitive to quantified manipulations (compared to direct measurement procedures that allow for resource-inefficient processing).[14]

Although the link to resource-efficiency seems to be the most promising route to exploring the quantification hypothesis with existing measurement procedures, some theorists have suggested that indirect measurement procedures in general can be reinterpreted from this perspective based on their demonstrated sensitivity to context. For example, Ferguson and Bargh (2007), arguing against the traditional assumption that behaviors can be explained as cognitive responses to discrete objects, have suggested that

[14] Resource-efficiency was selected as the feature of automaticity that best approximates the online/offline distinction not only for the positive reasons discussed here, but also for the negative reason that the other commonly proposed features of automaticity (Bargh, 1994) crosscut the online/offline distinction. In particular, offline processing is not assumed to be uniquely influenced by intentions to start or stop processing; instead, differences should depend on whether or not these intentions are quantified. Likewise, online processing is not assumed to operate outside conscious awareness per se, but is better understood as "first-order" awareness (see §4.2.3 for further discussion of this issue).

indirect measurement procedures reflect responses to "object-centered contexts." Likewise, De Houwer (2005) has suggested that the most commonly used indirect measurement procedure, the IAT, should be regarded as a general measure of *similarity*, which is essentially a function of the situational meaning of a stimulus. Although such reinterpretations of existing measurement procedures offer important challenges to the received view of dual-process cognition, the preceding discussion suggests that a note of caution is in order. Specifically, the reinterpretation of existing measurement procedures designed with the automatic/controlled distinction in mind may not be fully justified if newly proposed operating principles are assumed to correspond to *different* operating conditions (e.g., online/offline). Thus, before reinterpreting the results of existing direct and indirect measurement procedures or employing them in the investigation of newly proposed distinctions between kinds of cognition, more research is needed to understand which operating conditions existing procedures establish (De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009). Based on similarities and differences in the observation conditions that existing measurement procedures establish, inferences can then be drawn about what, exactly, the various existing procedures are measuring.

4.2.2 Implications for the development of new measurement procedures

4.2.2.1 Re-specifying the psychometric covariation thesis

Given the complications involved with reinterpreting existing measurement procedures, it is useful to consider how new measurement procedures might be designed to take the distinction between online and offline operating conditions into account. The first step is to draw out the implications of the quantification hypothesis for the psychometric covariation thesis, which requires identifying the observation conditions that correspond to online and offline operating conditions. In other words, what are the observation conditions that a measurement procedure must establish to select, as exclusively as possible, online or offline processing (and hence, via the cognitive covariation thesis, holistic or symbolic processing)?

As suggested above, the key distinction between online and offline operating conditions is the degree of sensory-dependence, which is inversely related to the capacity for metacognition: Sensory-dependence decreases as holistic representations are quantified, or “thought about,” as higher-order symbols. Consequently, online measurement procedures should establish observation conditions that preclude metacognition, whereas offline measurement procedures should facilitate the operation of this capacity. Given that metacognition requires effort and time, online measurement procedures should generally feature speeded responses to tap the early cognitive response to a stimulus and, as discussed in the previous section, should generally be unaffected by demands on cognitive resources. In addition, based on the assumption that holistic processing is distributed and hence non-linguistic, online measurement procedures should require responses that do not depend on linguistic processing. For the same reason, such procedures should not require discrete responses of any sort, but should make use of continuous measures that do not require the participant to discretize responses.

Offline measurement procedures would be characterized conversely, being designed to select conditions that facilitate the operation of metacognitive processing. Thus, these procedures should not require speeded responses, should not be employed under high resource-demand, should require linguistic, or at least verbalizable, responses, and should generally require discrete rather than continuous responses. The specification of online and offline measurement procedures in terms of the psychometric covariation thesis is presented in Table 4.2.

The Psychometric Covariation Thesis per the Quantification Hypothesis	
Observation Conditions	
Online	Offline
Speeded/Early	Non-speeded/Late
High resource demand	Low resource demand
Non-linguistic	Linguistic
Continuous	Discrete
Operating Conditions	
Online	Offline
High sensory-dependence (“Experiential”)	Low sensory-dependence (“Metacognitive”)

Table 4.2. The specification of the psychometric covariation thesis per the quantification hypothesis.

Whereas existing indirect measurement procedures designed to select implicit cognition often require speeded responses and are frequently employed under conditions of high resource demand, they do not, as a rule, require non-linguistic and continuous responses. These differences from the proposed characteristics of online measurement procedures may account for the lack of process purity for which indirect procedures are often criticized (see §1.1.1), and the relation between indirect and online measurement procedures is an important direction for future research. On the other hand, there is a much closer alignment between the characteristics of direct measurement procedures and those proposed for offline procedures. Generally, both direct and offline measurement procedures rely upon self-report, which is most often assessed without time constraints, under low resource demands, and which requires linguistic processing and discrete responses.

4.2.2.2 Measuring continuous cognition

Perhaps the most challenging aspect of online measurement, as described here, is the use of continuous rather than discrete responses. Continuous responses are assumed to better reflect holistic processing, which is not obviously segmented—either temporally or semantically—in the way that symbolic processing is (due to the dependence of symbolic processing on localist, compositional representation). It might be objected, however, that the notion of continuous measurement is a contradiction in terms, given that measurement inherently requires the discretization of observed behavior by placing it on a metric.[15] For just this reason, James (1890) saw little hope for the empirical study of continuous, holistic cognition:

> *The object of every thought … is neither more nor less than all that the thought thinks, exactly as the thought thinks it, however complicated the matter…. It is needless to say that memory can seldom accurately reproduce such an object, when once it has passed from before the mind. It either makes too little or too much of it. Its best plan is to repeat the verbal sentence, if there was one, in which the object was expressed. But for inarticulate thoughts there is not even this resource, and introspection must confess that the task exceeds her powers. The mass of our thinking vanishes for ever, beyond hope of recovery, and psychology only gathers up a few of the crumbs that fall from the feast. (p. 276)*

While it is true that measurement requires discretization, James perhaps did not place enough stock in the development of alternatives to direct/offline measurement procedures, which would make possible the observation of behavior at a much more fine-grained resolution. The outcomes of such procedures, being necessarily discretized at some level, are inevitably approximations of holistic processing, but these procedures nevertheless represent the best option for countering James' pessimism about the limits of empirical psychology.

---

[15] Indeed, this point suggests a powerful analogy between conceptions of science and reality on the one hand and symbolic and holistic cognition on the other: Each relation essentially involves measurement. Thus, the process of quantification, through which holistic representations are translated into symbolic representations, can be understood as a kind of introspective (metacognitive) measurement procedure. Taking the analogy a step further, symbolic representations might be viewed as abstract reconstructions of, or theories about, holistic representations.

In fact, measurement procedures have already been developed that possess many of the features of online measurement described above. For example, using a mouse-tracking procedure that continuously records the motion of a participant's hand during the generation of response behaviors, Freeman and Ambady (2009) found that early responses to stimuli are more holistic, being sensitive to features of the individual stimulus, before resolving to more discrete, categorical assessments. Spivey and Dale (2006) have reached similar conclusions using eye-tracking procedures. Wojnowicz, Ferguson, Dale, and Spivey (2009) also employed a mouse-tracking procedure to shed light on the cognitive processing underlying self-reported evaluations of social targets; their results suggested that early representations are sensitive to multiple sources of information before "self-organizing" into the categorical representations required by self-report procedures. In addition, a small but growing number of researchers have begun to apply the dynamical systems approach to psychology (e.g., Bechtel, 2008; Vallacher, Read, & Nowak, 2002; Wiese, Vallacher, & Strawinska, 2010). This approach, inspired by the cybernetics movement that originated in the mid-twentieth century, assumes that much cognitive processing is continuous and non-linear. A common feature of these various approaches to measurement is the emphasis on the time course of cognitive processing, which is a crucial aspect of the relation between holistic and symbolic cognition. In contrast, direct and indirect measurement procedures have typically ignored the temporal aspect of cognitive processing, providing only static cross-sections of presumably dynamic processes.

James' worries about holistic cognition touch upon a second problem for empirical psychology, however: Holistic cognition is not only temporally continuous, but semantically continuous as well. This is a consequence of the fundamental context-dependence of holistic representation, and it poses a serious threat to the practice of empirical psychology (Fodor, 1987). In short, if holistic cognition is tightly bound to the situational context in which it operates, then it becomes extremely difficult to produce valid generalizations about these psychological processes. The solution in mainstream psychology has been to ignore this possibility and assume that all cognition can be described in general, aggregate-level terms, essentially as if all cognition were symbolic (i.e., discrete, linguistic, etc.). In practice, this requires heavy dependence upon the use of ceteris paribus clauses in psychological explanation to subsume all of the unidentified variation that can influence the link between an "operationally defined" stimulus and a behavioral response (Fodor, 1989). This solution, however, seriously undermines the value of general psychological explanations. The difficulty is that psychologists seeking to provide *precise* descriptions of psychological mechanisms cannot ignore the possibility that some of those mechanisms may be context-dependent, and in these cases, sweeping the "random error" due to context-sensitivity into a ceteris paribus clause precludes the possibility of precisely describing the cognitive mechanism underlying behavior (Borsboom, Mellenbergh, & van Heerden, 2003; Cronbach, 1975; Molenaar, 2004; Runkel, 1990).[16]

The covariation theses offer the framework for a more adequate solution to this problem by helping researchers empirically distinguish between conditions under which cognitive processing will be more or less context-dependent. The quantification hypothesis thus suggests that under online (compared to offline) conditions, descriptions of cognitive processes will need to be more context-dependent. The practical problem with this suggestion is that mainstream empirical psychology depends heavily upon aggregation in the analysis of behavioral data (Epstein, 1979, 1980), and aggregation is inherently decontextualizing. One potential solution to this problem is the use of multi-level analyses, which allow for the modeling of data at both the aggregate and the individual level. Although the practice is not yet mainstream, the tools for measuring and analyzing multi-level data are already available (e.g., Hamaker, Dolan, & Molenaar, 2005; Molenaar, 2007; Nesselroade & Ram, 2004). Collectively, then, psychometric advances offer reason to doubt James' pessimistic conclusion: The measurement and analysis of online, holistic cognition is not only possible but empirically practical.

---

[16] In considering the threat holistic processing poses to the generality of psychological explanation, Conrey and Smith (2007) claim that "Fortunately, the aspects of context that are most important to social psychology are relatively few" (p. 254). To the extent this is true, the problem of distinguishing between behaviors that require context-dependent vs. context-independent explanations would be relatively innocuous, and traditional (aggregative) empirical methods could be applied unconditionally. In light of the wide variety of contextual cues to which social behavior is sensitive (see §1.2.1.2), however, the assumption that holistic processing is not *too* holistic appears unwarranted.

4.2.3 The possibility of graded distinctions

Beyond the measurement of continuous cognition, the quantification hypothesis presents another challenge for empirical research: Whether the distinction between kinds of cognition is *itself* discrete or continuous—and in the latter case, if the principle of covariation loses its value as a guide to theory and measurement (cf. §1.1.2). To begin with, the quantified relation between holistic tokens and symbolic types has been described as metacognitive. Yet metacognition implies a continuum, from first-order cognition through successively higher orders of cognition, each of which "objectifies" the preceding level (Rosenthal, 2002). If the transition from holistic to symbolic cognition were perfectly graded in this way, lacking any sharp distinction, the cognitive covariation thesis would indeed lose much of its value: It is covariation between *kinds* of operating principles and operating conditions that makes the thesis useful for drawing inferences about what kind of cognition has been measured and for designing measurement procedures to select specific kinds of cognition (Figure 1.2). In short, a graded distinction between kinds of cognition undercuts the process purity assumption that makes the principle of covariation useful for guiding empirical research. Without an empirical criterion for inferring when psychological mechanisms will be more or less context-sensitive, researchers will have no principled way to move beyond imprecise ceteris paribus explanations, as discussed above.

Of course, whether the distinction between kinds of cognition is qualitative or graded depends on assumptions about the operating principles that characterize them. The validation hypothesis, with its clear-cut distinction between sensitivity to perceived validity, is not amenable to graded distinctions. But the contextualization hypothesis, which distinguishes between relative sensitivity to situational context, could be interpreted in this way. In fact, Cunningham, Zelazo, Packer, and Van Bavel (2007) developed their Iterative Reprocessing Model specifically along these lines, such that the time course of cognitive processing upon observation of a stimulus is described as a loop in which each iteration adds contextual information to an initially rough object representation. The quantification hypothesis takes just the opposite approach, assuming that the time course of cognitive processing is essentially a process of decontextualizing initially holistic representations of the immediate situation. But the notion of iterative reprocessing still applies: Following the initial transition from first-order, holistic, to second-order, symbolic cognition, the “metacognitive loop” may continue to cycle, producing successively abstract, generalized representations with each iteration. An eventual product of these higher-level iterations may be communicable linguistic concepts.

The critical point, however, is that the initial iteration of this metacognitive loop, which is responsible for the translation of holistic representations to symbolic representations, involves a difference of *kind*, not just degree. As argued in §4.1.1, the quantification hypothesis is rooted in the assumption that there are two genuinely different kinds of mental representation (i.e., holistic and symbolic) that are not reducible. Thus, there is a qualitative difference between holistic, non-compositional representation and symbolic, compositional representation, analogous to the difference between a bitmap image of a scene and a linguistic description of the same scene—the description, but not the image, is built from *meaningful parts* (Fodor, 2007). Based on the assumptions specified in the quantification hypothesis (Table 4.1), therefore, the distinction between holistic and symbolic processing cannot be understood as thoroughly graded.

Assuming that the distinction between operating principles is qualitative implies that the distinction between operating conditions must be as well. With regard to the quantification hypothesis, this is true insofar as the operating conditions are understood as “experiential” (first-order) and “metacognitive” (higher-order), but the notion of sensory-dependence underlying these descriptions is naturally graded (see §4.1.1). This discontinuity, however, is an inevitable fact of empirical research. Operating conditions, being the link between the cognitive covariation thesis and the psychometric covariation thesis, serve a critical mediating role in the relation between theory (i.e., operating principles) and observation (i.e., observation conditions; Figure 1.2). In the case of the quantification hypothesis, the mapping between theory and observation is necessarily approximate, since the distinction between operating principles is assumed to be qualitative, but two of the four observation conditions specified in the psychometric covariation thesis (i.e., the speed and resource-dependence of responses) are graded

(Table 4.2). While measurement procedures either do or do not require linguistic or otherwise discrete responses, the speed and resource-dependence of those responses is more obviously a matter of degree. Thus, bringing the quantification hypothesis—and the underlying assumption of process purity—to bear upon empirical observation requires a bit of compromise. For this reason, it is recommended that measurement procedures designed or interpreted from the perspective of the quantification hypothesis be characterized in terms of the *relative* speed and resource-dependence required of responses (Keren & Schul, 2009; Moors & De Houwer, 2006). Establishing a useful standard of comparison for relative characterizations of observation conditions remains an important goal for future research.

4.2.4 Summary

The cognitive covariation thesis is an indispensible tool for resolving the problem of measurement in cognitivist psychology: It not only sheds light on what kind of process is being measured, but also guides the development of new measurement procedures based on assumptions about how cognitive processes operate. From the perspective of the quantification hypothesis, existing indirect and direct measurement procedures are not designed to select the appropriate operating conditions of cognition, and therefore cannot be straightforwardly reinterpreted in terms of holistic and symbolic processing. Nevertheless, evidence from these procedures may be informative to the degree that the indirect/direct and online/offline distinctions overlap, which is an important empirical question. Designing measurement procedures specifically based on the quantification hypothesis, however, raises the problem of measuring holistic cognition, which is assumed to be temporally and semantically continuous. Fortunately, psychometric advances suggest that the measurement of continuous cognition is not an insurmountable problem. Finally, because the distinction between holistic and symbolic cognition is qualitative rather than graded, the quantification hypothesis and the underlying principle of covariation remain useful as a guide to theory and research. The quantification hypothesis suggests that the metaphor of "carving cognition at its joints" is accurate, but is necessarily approximated in empirical applications given that some observation conditions are graded.

4.3 Future directions

Although the research reported in §2 and §3 is consistent with the quantification hypothesis, it represents only preliminary support for its central predictions. In future work, it will be critical to pit the key predictions of the contextualization, validation, and quantification hypotheses against one another. In addition, interpretation of this work should be sensitive to potential discrepancies between existing indirect measurement procedures and the proposed design of online measurement procedures. In this section, I will discuss four directions for future research suggested by the quantification hypothesis: first, an empirical approach to comparing the quantification and validation hypotheses; second, implications regarding the processing of negations; third, predictions regarding mediation patterns; and finally, implications for the interpretation and design of measurement procedures.

4.3.1 Comparing the quantification and validation hypotheses

As discussed in §4.1.2.1, the quantification hypothesis suggests that asymmetric influences on higher-level and lower-level processing are not due to *what* is manipulated but *how*; in particular, the hypothesis suggests that it is not the content of a manipulation but its quantification that should determine which kind of process it directly influences. This point can be used to draw out an important contrast with the validation hypothesis, which essentially picks out sensitivity to a particular contextual factor (i.e., situational consistency) as the key distinction between operating principles of cognition. The basic approach would involve crossing how the meaning of a stimulus is manipulated (i.e., via reassignment vs. change in situational context) with the content of the manipulation—in this case, the perceived validity of a stimulus. Practically speaking, this would require manipulating perceived validity either through direct assignment (i.e., labeling stimuli as true or false) or through the situational context (i.e., such that the truth or falsity of a stimulus is implied by the situation in which it is perceived). Whereas the validation hypothesis would predict a holistic/symbolic dissociation due to content (i.e., the perceived validity of the stimulus), the quantification hypothesis would predict a dissociation due to how perceived validity was manipulated (i.e., via direct assignment vs. situational context).[17]

Interestingly, the existing social-cognitive literature relevant to these predictions is mixed. Gawronski and Strack's (2004) finding that induced compliance manipulations influence explicit but not implicit evaluations (see §1.2.2.2) appears to disconfirm the prediction of the quantification hypothesis, assuming that the manipulation altered the context in which the target stimulus was perceived rather than directly reassigning its truth-value. On the other hand, Schul, Mayo, and Burnstein (2004) found that statements delivered by untrustworthy-looking faces were automatically perceived to be false; in this case, the contextual manipulation of perceived validity did appear to have a direct influence on lower-level processing. Future work might seek to disentangle these findings, perhaps by focusing on differences in how manipulations of situational context are empirically realized.

4.3.2 Negation

The quantification hypothesis suggests that in Experiment 3.3, it was not temporal delay that caused the dissociation between implicit and explicit evaluations but rather the different ways in which object-valence associations were negated (see also Gregg et al., 2006). In the short-delay condition (similar to Experiments 3.1 and 3.2), participants were able to negate false associations immediately after viewing them, which amounted to token-level, non-quantified negation. In the long-delay condition, however, the negation had to be applied to an entire set of false associations, which amounted to type-level,

---

[17] This approach would be less useful for empirically comparing the quantification and contextualization hypotheses, since they essentially differ with regard to the influence of *how* the meaning of a stimulus is manipulated without making any claims about the differential influence of particular kinds of content (e.g., perceived validity). It is in this sense that the quantification hypothesis is the converse of the contextualization hypothesis: The former predicts that manipulating the meaning of a stimulus via changes in situational context should directly influence lower-level processing, whereas the latter predicts a direct influence on higher-level processing. This is because, from the perspective of the contextualization hypothesis, controlled processing is assumed to make possible the incorporation of situational context for individuating initially stereotypical representations. Notably, this prediction has been repeatedly disconfirmed by evidence that manipulations of situational context directly influence implicit processing (see §1.2.1.2 and Experiment 2.1).

quantified negation. As argued in §4.1, there is no obvious reason to expect that temporal delay *per se* should lead to a primacy effect on implicit but not explicit evaluations. Thus, an important direction for future research is to deconfound time and quantification as potential moderators of the influence of negation on implicit (or online) evaluations.

On the other hand, the quantification hypothesis is challenged by evidence that non-quantified, token-level negations fail to influence implicit evaluations (e.g., Deutsch, Gawronski, & Strack, 2006). Although the influence of negation on implicit evaluations appears to depend on the type of indirect measurement procedure used (Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009), at least when using an evaluative priming task the negation of valenced primes is not efficiently processed unless the negation has a familiar referent. For example, "no cockroach" fails to facilitate evaluations of positive target words, but "no luck" does facilitate evaluations of negative target words (see also Mayo, Schul, & Bernstein, 2004). At a conceptual level, this finding is actually consistent with the quantification hypothesis because a negation that lacks a concrete (token-level) referent (e.g., "no cockroach") is inherently an abstract concept, and therefore should be incapable of holistic representation according to the principles proposed earlier (Table 4.1). As Osherson and Smith (1981) pointed out, the "fuzzy logic" by which distributed connectionist networks operate is incapable of representing logical truth-values *per se*, but instead must represent them in terms of concrete instances; without a concrete referent, a negation is purely an abstract concept. A second response to these findings is that, because Deutsch et al.'s (2006) evaluative priming task used linguistic negations, the procedure was not actually measuring online (and hence holistic, token-level) cognition (Table 4.2). Thus, an interesting prediction of the quantification hypothesis would be that, using an *online* measurement procedure, pictorial negations should be processed efficiently only when they have a concrete referent, but linguistic negations should not be processed even when a concrete referent is available.

4.3.3 Mediation patterns

The quantification hypothesis also makes specific predictions regarding the mediation patterns that should be observed between holistic and symbolic cognition, due to differences in how each is directly influenced and the nature of bottom-up (inductive) and top-down (deductive) processing. Bottom-up mediation patterns should be observed when the meaning of a stimulus is manipulated via changes in the situational context that are explicitly unidentified. Such manipulations should still have a direct influence on holistic processing, but would influence symbolic processing only indirectly via metacognitive awareness. It remains an interesting question whether changes in the situational context that are explicitly identified nevertheless produce simultaneous influences on holistic and symbolic cognition, or if explicit identification of contextual manipulations somehow disrupts their influence on holistic meaning. Top-down mediation patterns should be observed when meaning is directly reassigned. Experiment 2.2 provides a relevant example. In this case, the newspaper clipping caused participants to directly assign new meanings to their self-concepts (e.g., "I am extraverted"). This reassignment then appeared to cause participants to engage in a confirmatory search through token-level self-knowledge stored in memory, resulting in a top-down influence on holistic cognition.

4.3.4 Measurement

The quantification hypothesis implies that current indirect measurement procedures are not optimized for the assessment of holistic cognition. An important goal for future research, then, is to design genuinely online measurement procedures and to investigate the extent to which existing indirect procedures might be considered online. As discussed in §4.2.2, there have already been advances in online measurement that make use of mouse-tracking and eye-tracking (e.g., Freeman & Ambady, 2009; Spivey & Dale, 2006; Wojnowicz et al., 2009), and these procedures can be employed to test the major predictions of the quantification hypothesis. For example, the quantification and contextualization hypotheses essentially disagree on how the time-course of mental representation is characterized: either from context-dependent to context-independent or the reverse. Online measurement procedures are necessary for distinguishing between early and late mental representations, and are therefore crucial for testing differences in context-sensitivity over time.

With regard to investigating the extent to which existing indirect measurement procedures can be considered online, research suggests that the most popular indirect procedure, the IAT, is especially sensitive to type-level as opposed to token-level knowledge (Nosek, Greenwald, & Banaji, 2005; Olson & Fazio, 2003). For example, Foroni and Bel-Bahar (2009) found that IAT effects were significantly larger when stimuli were words rather than pictures, presumably because there was a closer match in level of representation between the abstract response categories and the words. Findings such as these suggest that the IAT may not be particularly adaptable to online measurement, as it relies heavily upon categorical (and hence symbolic) knowledge.

Other indirect measurement procedures, however, may be more suitable given that they are less reliant upon categorical responses. For example, preliminary data from a study I have conducted with Bertram Gawronski suggest that the AMP, with slight modifications, can provide a window onto holistic processing. In earlier work, Gawronski, Cunningham, LeBel, and Deutsch (2010) asked participants to pay attention to either the race (black vs. white) or the age (young vs. old) of the primes in an AMP and an EPT. They found that implicit evaluations measured with the EPT were not influenced by the unattended dimension but that responses on the AMP were, suggesting that primes in the AMP were perceived more holistically. The present study sought to test the prediction, derived from the quantification hypothesis, that the context-sensitivity of responses on the AMP would be higher early on and would decrease over time, once the stimuli became represented in terms of discrete categories. In order to test this prediction, the delay between presentation of the prime and the Chinese character was manipulated, such that in the "early response" condition the delay was 125 ms (the standard stimulus-onset asynchrony used for the procedure) whereas in the "late response" condition the delay was 1000 ms (see Hofmann, van Koningsbruggen, Stroebe, Ramanathan, & Aarts, 2010, for a similar modification of the AMP). Although the data pattern was not statistically significant, the trends were consistent with the prediction that responses made quickly after viewing a stimulus were influenced by both race and age, whereas responses made relatively later were influenced more strongly by the attended category. These findings suggest that there may be substantial overlap between particular indirect measurement procedures and the proposed design of online procedures, but this remains an empirical question that will benefit from future research.

4.4 Conclusion: Placing the quantification hypothesis in context

In closing, I would like to address a much broader question: Is the quantification hypothesis supposed to provide the basis for all psychological explanations of behavior? The short answer is no; the hypothesis is meant to provide structure specifically for *intentional* explanations, which appeal to an individual's mental representations (i.e., what a person is thinking about) to explain behavior (Dennett, 1987; Haugeland, 1978).[18] One way to view the contribution of the quantification hypothesis is that it attempts to structure intentional explanations when the possibility of non-localist, non-compositional representation is taken into account. The intentional explanations at which the quantification hypothesis is aimed characterize the cognitivist, subjective tradition of empirical psychology. There is, however, a behaviorist, objective tradition as well, which regards mental representation as externally determined (De Houwer, 2011; Fodor, 1980; Putnam, 1975). Psychological explanations formulated from this latter perspective are effectively non-intentional, since any internal variance in mental representation is assumed not to play an explanatory role in the stimulus-response link. In its purest form, this approach assumes that behavior can be explained in terms of veridical perception of the environment.

Intriguingly, the cognitivist and behaviorist approaches to psychological explanation each appear to be valid, though under different conditions. For example, it has been found that perceptions of the size of objects are influenced by cognitive illusions, but that motor movements designed to interact with those objects are nevertheless well calibrated to their actual size (e.g., Aglioti, DeSouza, & Goodale, 1995; Goodale & Humphreys, 1998). Such findings suggest that a higher-order covariation thesis may be required to distinguish between the conditions under which behavior is better explained with cognitivist vs. behaviorist principles. In specifying the quantification hypothesis, I have

---

[18] In Franz Brentano's original formulation, intentionality is the mark of the mental; in other words, the capacity to represent, or be "about," something else, is what distinguishes mental from physical phenomena. The present usage of the term *intentional* should not be confused with its more colloquial meaning, which typically denotes a planned behavior.

assumed that the world does not arrive at perception pre-categorized, and hence that quantification (i.e., the categorization of percepts) is an effortful process. But there may be senses in which the world can be viewed as pre-categorized, most obviously with regard to “natural” or biological kinds (e.g., Bird & Tobin, 2009; Millikan, 1984), and it is reasonable to think that cognition would have evolved to exploit these regularities—even though they are ultimately no less theoretical than ad hoc categories. Thus, the possibility of operating conditions that distinguish between cognitivist and behaviorist principles of cognition, and their potential interaction (i.e., the point of connection between “narrow,” subjective and “wide,” objective representation; Loar, 1988; Pereboom, 1995) offer deeply interesting questions for future empirical and theoretical work.

4.5 References

Aglioti, S., DeSouza, J.F.X., Goodale, M.A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*, 679–685.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, *105*, 181–199.

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

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230-244.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bird, A., & Tobin, E. (2009). Natural kinds. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2009 ed.). Retrieved from http://plato.stanford.edu/archives/spr2009/entries/natural-kinds/

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203-219.

Cesario, J., Plaks, J. E., & Higgins, E. T. (2006). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social Psychology*, *90*, 893-910.

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (53-115). Oxford: Oxford University Press.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*, 736-760.

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

De Houwer, J., Geldof, T., & De Bruycker, E. (2005). The Implicit Association Test as a general measure of similarity. *Canadian Journal of Experimental Psychology*, *59*, 228-239.

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

Dennett, D. C. (1987). *The intentional stance*. Cambridge: The MIT Press.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385-405.

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

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735-755.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*, 1097-1126.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist*, *35*, 790-806.

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297-327.

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 216-246). New York: Guilford.

Fodor, J. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, *3*, 63-73.

Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge: The MIT Press.

Fodor, J. (1989). Making mind matter more. *Philosophical Topics*, *17*, 59-79.

Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105-116). Malden, MA: Blackwell.

Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Cambridge, MA: Blackwell.

Foroni, F., & Bel-Bahar, T. (2010). Picture-IAT versus Word-IAT: Level of stimulus representation influences on the IAT. *European Journal of Social Psychology*, *40*, 321-337.

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single reading of a short scenario. *Psychonomic Bulletin & Review*, *12*, 139-144.

Freeman, J. B., & Ambady, N. A. (2009). Motions of the hand expose partial and parallel activation of stereotypes. *Psychological Science*, *20*, 1183-1188.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731.

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

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

Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67*, 181-207.

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

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

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207-233.

Haugeland, J. (1978). The nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, *1*, 215-260.

Hofmann, W., van Koningsbruggen, G. M., Stroebe, W., Ramanathan, S., & Aarts, H. (2010). As pleasure unfolds: Hedonic responses to tempting food. *Psychological Science*, *21*, 1863-1870.

Horst, S. (2009). The computational theory of mind. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2009 ed.). Retrieved from http://plato.stanford.edu/archives/win2009/entries/computational-mind/

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Henry Holt & Co.

Keren, G., & Schul, Y. (2009) Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533-550.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

Lee, S. W. S., & Schwarz, N. (2010). Washing away postdecisional dissonance. *Science*, *328*, 709.

Loar, B. (1988). Social content and psychological content. In R. H. Grimm & D. D. Merrill (Eds.), *Contents of thought* (pp. 99-110). Tucson, AZ: The University of Arizona Press.

Lord, C. H., & Lepper, M. R. (1999). Attitude representation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 265-343). San Diego, CA: Academic Press.

Markus, H. R. (2005). On telling less than we can know: The too tacit wisdom of social psychology. *Psychological Inquiry*, *16*, 180-184.

Mayo, R., Schul, Y., & Burnstein, E. (2004). “I am not guilty” vs. “I am innocent”: Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*, 433-449.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge: The MIT Press.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201-218.

Molenaar, P. C. M. (2007). Psychological methodology will change profoundly due to the necessity to focus on intra-individual variation. *Integrative Psychological and Behavioral Science*, *41*, 35-40.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297-326.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9-29.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135-183.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Emobdiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*, 184-211.

Nisbett, R., E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, *108*, 291-310.

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

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166–180.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35-58.

Pereboom, D. (1995). Conceptual structure and the individuation of content. *Philosophical Perspectives*, *9*, 401-426.

Perugini, M., & Prestwich, A. (2007). The gatekeeper: Individual differences are key in the chain from perception to behavior. *European Journal of Personality*, *21*, 303-317.

Putnam, H. (1975). The meaning of “meaning”. *Minnesota Studies in the Philosophy of Science*, *7*, 131-193.

Rey, G. (2007). Resisting normativism in psychology. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105-116). Malden, MA: Blackwell.

Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 1-59). Mahwah, NJ: Erlbaum.

Rosenthal, D. M. (2002). Explaining consciousness. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 406-421). Oxford: Oxford University Press.

Runkel, P. J. (1990). *Casting nets and testing specimens: Two grand methods of psychology*. New York: Praeger.

Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, *86*, 668-679.

Schwarz, N. & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513-523.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology*, *89*, 583-592.

Smith, E. R., & Semin, G. R. (2004). Socially situated cognition: Cognition in its social context. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 53-117). San Diego: Elsevier.

Smith, J. D., Ashby, F. G, Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin and Review*, *18*, 414-421.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, *133*, 398-414.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*, 207-211.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: The University of Chicago Press.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*, 403-421.

Vallacher, R. R., Read, S. J., & Nowak, A. (2002). The dynamical perspective in personality and social psychology. *Personality and Social Psychology Review*, *6*, 264-273.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*, 3-15.

Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The active-self account. *Personality and Social Psychology Review*, *11*, 234-261.

Wiese, S. L., Vallacher, R. R., & Strawinska, U. (2010). Dynamical social psychology: Complexity and coherence in human experience. *Social and Personality Psychology Compass*, *4*, 1018-1030.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*, 606-607.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*, 625-636.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345-411). New York: Elsevier.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*, 1428-1435.

Appendix A:

Documentation for Ethics Approval

Western

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	07 10 21	**Approval Date**	07 10 31
Principal Investigator	Bertram Gawronski/Kurt Peters	**End Date**	08 04 30
Protocol Title	How do we think of our own personality		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2007-2008 PREB are: Mike Atkinson, David Dozois, Bill Fisher and Matthew Maxwell-Smith

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

[redacted]

Western

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	08 06 06	**Approval Date**	08 06 26
Principal Investigator	Bertram Gawonski	**End Date**	08 09 30
Protocol Title	Biographical memory and perceptions of media reports		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2007-2008 PREB are: Mike Atkinson, David Dozois, Bill Fisher and Matthew Maxwell-Smith

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	09 06 01	**Approval Date**	09 06 08
Principal Investigator	Bertram Gawronski/Kurt Peters	**End Date**	09 09 30
Protocol Title	Does attention influence attitudes?		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2008-2009 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Western

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	09 09 19	**Approval Date**	09 09 24
Principal Investigator	Bertram Gawronski/Kurt Peters/Yang Ye	**End Date**	10 04 30
Protocol Title	How do we form first impressions and evaluations?		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2008-2009 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Western

Use of Human Subjects - Ethics Approval Notice

Review Number	09 12 07	**Approval Date**	09 12 15
Principal Investigator	Bertram Gawronski/Kurt Peters/Amy Tapley	**End Date**	10 06 01
Protocol Title	How do we comprehend reports about scientific findings?		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2009-2010 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	10 05 04	**Approval Date**	10 05 12
Principal Investigator	Bertram Gawronski/Kurt Peters	**End Date**	10 09 30
Protocol Title	How doe we comprehend reports about scientific findings?		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2009-2010 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Western

Department of Psychology The University of Western Ontario
Room 7418 Social Sciences Centre,
London, ON, Canada N6A 5C1
Telephone: (519) 661-2067Fax: (519) 661-3961

Use of Human Subjects - Ethics Approval Notice

Review Number	10 10 01	**Approval Date**	10 10 01
Principal Investigator	Bertram Gawronski/Kurt Peters	**End Date**	11 04 30
Protocol Title	**How does your personality influence your thoughts?**		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

[redacted]

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2009-2010 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Appendix B:

Permissions from Publishers

Permission for §2

Request

ELSEVIER LICENSE
TERMS AND CONDITIONS
Apr 04, 2011
This is a License Agreement between Kurt R Peters ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.
All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.
Supplier
Elsevier Limited
The Boulevard,Langford Lane
Kidlington,Oxford,OX5 1GB,UK
Registered Company Number
1982084
Customer name
Kurt R Peters
Customer address
Dept. of Psychology, SSC
London, ON N6A 5C2
License number
2600900435176
License date
Feb 02, 2011
Licensed content publisher
Elsevier
Licensed content publication
Journal of Experimental Social Psychology
Licensed content title
Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning
Licensed content author
Kurt R. Peters, Bertram Gawronski
Licensed content date
4 December 2010
Licensed content volume number
n/a
Licensed content issue number
n/a
Number of pages
1
Start Page
End Page
Type of Use
reuse in a thesis/dissertation
Portion
full article

Format
both print and electronic
Are you the author of this Elsevier article?
Yes
Will you be translating?
No
Order reference number
Title of your thesis/dissertation
Carving Cognition at its Joints: Insights from the Interaction between Explicit and Implicit Cognition
Expected completion date
May 2011
Estimated size (number of pages)
200
Elsevier VAT number
GB 494 6272 12
Permissions price
0.00 USD
Value added tax 0.0%
0.00 USD / GBP
Total
0.00 USD

Permission

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
17. **Author website** for journals with the following additional clauses:
All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download

and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com
All content posted to the web site must maintain the copyright information line on the bottom of each image
You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com
20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.
21. **Other Conditions**:
v1.6

Permission for §3

Copyright Clearance Center RightsLink®

Home | Account Info | Help

SAGE

Title: Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations

Author: Kurt R. Peters, Bertram Gawronski

Publication: Personality & Social Psychology Bulletin

Publisher: Sage Publications

Date: 03/24/2011

Logged in as:
Kurt Peters
Account #:

LOGOUT

Redirected Request

If you are an Author inquiring about the re-use of your journal article, please note that after publication of the journal article, Authors may re-use their content in any print-form work written or edited by the Author or for classroom use, without seeking permission from SAGE. You may post an abstract of your journal article on the internet at any time. For any other use of your work, please contact the publisher. For additional information, see www.sagepub.com/permissions/nav.

BACK | CLOSE WINDOW

Curriculum Vitae

Name:	Kurt R. Peters
Post-secondary Education and Degrees:	The University of Western Ontario London, Ontario, Canada 2007-2011 Ph.D., Psychology
	The University of Western Ontario London, Ontario, Canada 2005-2007 M.Sc., Psychology
	Dartmouth College Hanover, New Hampshire, USA 1999-2003 B.A., Computer Science & Music
Honours and Awards:	Province of Ontario Graduate Scholarship 2008-2009, 2009-2010, 2010-2011
	Travel Award, Society for Personality and Social Psychology 2007
Related Work Experience	Teaching Assistant The University of Western Ontario 2005-2011

Publications:

Peters, K. R., & Gawronski, B. (2011). Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology*, *47,* 436-442.

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

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

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24*, 218-225.

Gawronski, B., Peters, K. R., Brochu, P., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, *34*, 648-665.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass*, *2*, 1002-1023.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives in Psychological Science*, *2*, 181-193.

Book Chapters:

Gawronski, B., Peters, K. R., & Strack., F. (2008). Cross-cultural differences in cognitive dissonance: A conceptual reanalysis. In R. M. Sorrentino, & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 297-314). New York: Elsevier.

Symposia:

Peters, K. R., & Gawronski, B. (2010, January). *What's so special about the self? Or is there anything special at all?* Symposium at the 11th Annual Meeting of the Society for Personality and Social Psychology (SPSP), Las Vegas, NV, USA. [Speakers: Kenneth DeMarree, Allen McConnell, Marret Noordewier, Kurt Peters]

Conference Presentations:

Peters, K. R., & Gawronski, B. (2011, January). *Are we puppets on a string? The origin of dissociations between implicit and explicit evaluations*. Poster presented at the 11th Annual Meeting of the Society for Personality and Social Psychologists (SPSP), San Antonio, TX, USA.

Peters, K. R., & Gawronski, B. (2010, October). *Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations*. Informal paper presented at the Annual Meeting of the Society of Experimental Social Psychology (SESP), Minneapolis, MN, USA.

Peters, K. R., & Gawronski, B. (2010, May). *Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations*. Paper presented at the Annual Western-Waterloo Meeting, London, ON, Canada.

Peters, K. R., & Gawronski, B. (2010, January). *An activation/validation model of self-construal*. Paper presented at the 11th Annual Meeting of the Society for Personality and Social Psychologists (SPSP), Las Vegas, NV, USA.

Peters, K. R., & Gawronski, B. (2009, May). *An activation/validation model of self-construal*. Poster presented at the Annual Western-Waterloo Meeting, Waterloo, ON, Canada.

Brochu, P. M., Gawronski, B., Peters, K. R., Strack, F., & Esses, V. M. (2009, June). *Cognitive consistency and the relation between different forms of prejudice: A review of the integrative prejudice framework*. Paper presented at Cognitive consistency as an integrative concept in social cognition. EASP Small Group Meeting, Bronnbach, Germany.

Peters, K. R., & Gawronski, B. (2009, May). *An activation/validation model of self-construal*. Paper presented at the 81st Annual Meeting of the Midwestern Psychological Association (MPA), Chicago, IL, USA.

Peters, K. R., & Gawronski, B. (2009, February). An *activation/validation model of self-construal*. Poster presented at the 10th Annual Meeting of the Society for Personality and Social Psychologists (SPSP), Tampa, FL, USA.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008, October). *Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms*. Informal paper presented at the Annual Conference of the Society of Experimental Social Psychology (SESP), Sacramento, CA, USA.

Peters, K. R., & Gawronski, B. (2008, May). *An activation/validation model of self-construal*. Paper presented at the Annual Western-Waterloo Meeting, London, ON, Canada.

Peters, K. R., Gawronski, B., & Deutsch, R. (2008, February). *Optimizing the reliability of the dot probe task as a measure of attentional bias: A time-course analysis*. Poster presented at the 9th Annual Meeting of the Society for Personality and Social Psychologists (SPSP), Albuquerque, NM, USA.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008, February). *Cognitive consistency and the relation between different kinds of prejudice: Implications for prejudice reduction*. Paper presented at the 9th Annual Meeting of the Society for Personality and Social Psychology (SPSP), Albuquerque, NM, USA.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008, February). *What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures*. Informal paper presented at the 8th Social

Cognition Preconference at the 9th Annual Meeting of the Society for Personality and Social Psychology (SPSP), Albuquerque, NM, USA.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2007, October). *Personal vs. extra-personal associations in implicit attitude measures: Conceptual issues in a methodological debate*. Informal paper presented at the Annual Conference of the Society of Experimental Social Psychology (SESP), Chicago, IL, USA.

Peters, K. R., & Gawronski, B. (2007, January). *I think I am, therefore I am: Exploring the mechanism of self-belief change through role-taking*. Poster presented at the 8th Annual Meeting of the Society for Personality and Social Psychologists (SPSP). Memphis, TN, USA.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2006, October). *What do implicit measures tell us? Scrutinizing the validity of three common assumptions*. Informal paper presented at the Annual Conference of the Society of Experimental Social Psychology (SESP), Philadelphia, PA, USA.

Peters, K. R., & Gawronski, B. (2006). *Associative and propositional processes in prejudice: Understanding the relation between different forms of racism*. Poster presented at the 7th Annual Meeting of the Society for Personality and Social Psychologists (SPSP). Palm Springs, CA, USA.