Implicit Cognition and Psychopathology: Looking Back and Looking Forward

Bethany A. Teachman,1 Elise M. Clerkin,2 William Cunningham,3 Sarah Dreyer-Oren,2 and Alexandra Werntz1

1 Department of Psychology, University of Virginia, Charlottesville, Virginia 22904, USA; email: bteachman@virginia.edu
2 Department of Psychology, Miami University, Oxford, Ohio 45056, USA
3 Department of Psychology, University of Toronto, Toronto, ON M5S 3G3, Canada

Keywords
implicit association, psychopathology, automatic, anxiety, depression, alcohol

Abstract
Implicit cognitive processing is theorized to have a central role in many forms of psychopathology. In the current review, we focus on implicit associations, by which we mean evaluative representations in memory that are difficult to control and do not require conscious reflection to influence affect, cognition, or behavior. We consider definitional and measurement challenges before examining recent empirical evidence for these associations in anxiety, obsessive–compulsive, posttraumatic stress, depressive, and alcohol use disorders. This examination is framed by a brief review of the ways that prominent models of psychopathology represent biased implicit processing of disorder-relevant information. We consider to what extent models reflect more traditional automatic/implicit versus strategic/explicit dual-process perspectives or reflect more recent dynamical systems perspectives in which mental representations are iteratively reprocessed, evolving continuously. Finally, we consider the future research needed to better understand the interactive and temporal dynamics of implicit cognition in psychopathology.
INTRODUCTION

There is mounting empirical and theoretical evidence supporting the notion that cognitive processing that occurs outside one's control, and sometimes even one's awareness, is a defining feature of many forms of psychopathology (e.g., Teachman et al. 2012). This relatively automatic processing helps to clarify why another otherwise rational person refuses to venture into her basement for fear that she will be bitten by a spider despite the fact that there are no venomous spiders in the part of the country where she lives. It also helps to explain why the doctor with alcohol use disorder (AUD) desperately wants a drink even though she knows her drinking is ruining her marriage and career, and why the recent college graduate with depression feels stuck in an endless cycle of demoralizing, self-critical thinking about his future job prospects. Given these responses, it is not surprising that researchers have written for several decades about the defining role of relatively automatic processing in psychopathology. As just one example, in McNally's seminal paper regarding automaticity and anxiety, he concluded, “It is the inability of the patient to terminate fear-generating processing once it starts that is the hallmark of pathological anxiety” (McNally 1995, p. 752).

In the current review, we discuss evidence for how individual differences in implicit cognition can inform our understanding of many common forms of psychopathology. We focus on so-called implicit associations, by which we mean representations in memory that link a stimulus and an involuntarily activated evaluative outcome and that do not require conscious reflection to influence affect, cognition, or behavior. Our goal is not to review every paradigm intended to measure the representations or every disorder affected by these associations—that would require a book—but to provide an overview of what we know about the part these associations play in the etiology, maintenance, and reduction of core clinical problems. With this knowledge, we then consider how advances in theory and empirical research can better position us to identify and manage the insidious effects of implicit associations in psychopathology in the future.
It is clear that there are reasonably reliable individual differences in how people associate different stimulus or context cues with expected evaluative outcomes. For example, someone who automatically evaluates spiders as more threatening than someone else does likely has a stronger memory association between spiders (the stimulus) and danger (a threat evaluative outcome). For historical reasons, and to link to the larger literature, we refer to these differences in memory as implicit associations, but we do not make claims regarding whether these representations in memory also inform more deliberative processes, and we do not assume distinct systems of implicit association and explicit association (see the discussion in Jones et al. 2014). Thus, although we use the term implicit associations, we wish to avoid some of the conceptual disagreements regarding the possible distinctions between implicit and explicit representations. Specifically, there is a debate regarding whether implicit evaluations are informed by a different set of memory representations than more explicit evaluations or whether both relatively more implicit and explicit evaluations can rely on the same set of representations (see the discussion in Jones et al. 2014).

Because we focus here specifically on implicit evaluations and measures, the specific nature of the relationship between implicit and explicit processes and its implications for clinical psychology are outside of the scope of this review. Rather, we use the terms implicit cognition and implicit associations throughout this review for their heuristic value, but we remain agnostic to the debates surrounding their computational processes. Instead, we simply acknowledge that there are representations in memory that can be quickly and unintentionally used to inform evaluative judgments and that do not require conscious reflection or control.

When measures of implicit associations were developed, the dominant models of psychology (e.g., Greenwald & Banaji 1995, Schneider & Shiffrin 1977) proposed two modes of cognition. One mode was the deliberative, strategic, and conscious sets of cognitive operations of which we are most aware. These were contrasted with a set of processes that were thought to require no intention to initiate, happened outside of conscious awareness, did not require cognitive resources, and were not controllable. Although different models gave these two sets of processes different labels (e.g., System 1 versus System 2; Kahneman 2003), the first set was often referred to as explicit cognition, whereas the second set was referred to as implicit cognition. Some models even suggested that implicit and explicit cognition were so dissociated that they used separate, non-interacting sets of representations (e.g., Greenwald & Banaji 1995). This meant that one could believe or feel one thing at one level of cognitive process, but believe or feel something quite different at another. In other words, one could know consciously that spiders were perfectly fine and even interesting, but still unconsciously react negatively. Because of these models, researchers often assumed that evidence for one processing feature implied each of the others in that mode. For example, if one found evidence that people could not control (i.e., not alter or stop) their evaluation of an ambiguous facial expression as threatening (typical of social anxiety), it was assumed that people were also unconscious (not consciously aware) of this evaluation and that it occurred unintentionally (without a goal to initiate it) and efficiently (consuming few-to-no processing resources). Unfortunately, this has resulted in a literature in which the term implicit can mean many different things. The term is often used interchangeably with automatic, and it is used as an umbrella concept that subsumes many different features that do not consistently covary (e.g., processing can be unconscious but not uncontrollable or unintentional; see Bargh 1994, Moors & De Houwer 2006).

Given the challenges in defining implicit associations, it is helpful to consider how recent social, cognitive, and neuropsychological models of implicit cognition move away from artificially separating automatic/implicit from strategic/explicit forms of processing and instead apply a dynamical systems perspective in which mental representations are evolving continuously as different partial representations compete and interact in an iterative way (Jones et al. 2014). For
example, according to the iterative reprocessing model (Cunningham & Zelazo 2007), the evaluative system is built upon multiple component processes that do not work in an all-or-none fashion. According to the model, stimulus representations (e.g., people, objects, or abstract concepts) initiate a sequence of processes through which information is interpreted and then repeatedly reinterpreted in light of an increasingly rich set of contextually meaningful representations. Whereas evaluations based on few iterations of the evaluative cycle are relatively automatic in that they are obligatory and may occur without conscious monitoring, evaluations based on additional iterations and computations are relatively reflective. In this system, lower-order evaluative processes continue to provide affectively laden information about valence (positive and negative) and arousal even as additional higher-order processes are recruited. With each additional iteration, information from all activated processes can be passed back to the relatively lower-order processes and the evaluation is recalculated: New attitude representations and contextual information can then be activated or foregrounded to help construct a more carefully considered evaluation.

With these potentially different patterns or combinations of activation, evaluations of the same information can result in distinct subjective evaluative experiences and, perhaps, even different evaluative outcomes at different times. Unlike a dual-process, automatic-then-strategic linear model, a dynamical systems perspective allows for continuous updating at all levels of processing.

While there continues to be considerable debate regarding the processes underlying implicit cognition, understanding the predictive validity and clinical utility of implicit association measures does not require the endorsement of one framework or another regarding whether and how implicit associations inform deliberative processing or when implicit processing features will be independent or co-occur. In this regard, this article reviews both the existing literature on implicit associations in psychopathology, which tends to use a more static measurement approach that looks at implicit associations in isolation, and also considers what is needed so that future research can more effectively study how the processing of disorder-relevant information is transformed over time.

It is worth noting that our emphasis on the key role played by automatic processes in psychopathology does not reflect a parallel devaluing of more strategic forms of processing. More controlled, intentional, and conscious processes are undoubtedly important in the expression of many, if not all, disorders. The woman with a spider phobia, for instance, may ultimately choose to venture into her basement if she consciously reflects on the fact that her likelihood of being bitten by a venomous spider is virtually nonexistent or if she overrides her unintentionally activated belief that she cannot cope around spiders. Further, as more research demonstrates that involuntary processing is not immune from strategic control (e.g., Cunningham et al. 2008), changing how one deliberates may, under some circumstances, be useful in modifying the type of automatic responses that people have. Thus, we highlight the evidence for implicit psychopathology associations with the understanding that it is actually the dynamic interaction across different automatic and strategic processes over time that is the key target to promote recovery.

**SCOPE OF THE REVIEW**

This article reviews how implicit cognitive processing has been understood and assessed in key areas of psychopathology. Our review focuses on anxiety (specific phobias, and panic, social anxiety, and generalized anxiety), and obsessive–compulsive, posttraumatic stress, depression, and alcohol use disorders because these are common areas of psychopathology that each has a substantial theoretical or empirical literature, or both, on implicit associations. We recognize that this disorder list is somewhat arbitrary and appreciate that other psychopathology domains, such as eating, psychotic, body dysmorphic, and personality disorders, as well as suicide, also have important implicit
association research. We decided that only a subset of the many possible disorder domains could adequately be considered given the space constraints (and because recent reviews have occurred in some of these other domains; e.g., eating disorders: Smith & Forrest 2017; suicide: Cha et al. 2018). Analogously, not every implicit association paradigm is reviewed, but the main ones used in the clinical literature are considered, including the Implicit Association Test (IAT), Single Category IAT, Brief IAT, Go/No-Go Association Task (GNAT), Implicit Relational Assessment Procedure (IRAP), and Evaluative (or Affective) and Semantic Priming. After describing these measures, we consider the prominent cognitive theories and the evidence from these implicit association measures for each disorder domain. Finally, we consider the barriers that need to be overcome to make more progress in this field and call for future research.

MEASURES TO ASSESS IMPLICIT ASSOCIATIONS IN PSYCHOPATHOLOGY

A variety of tasks have been used to reflect implicit psychopathology associations, with the IAT (Greenwald et al. 1998) used most widely. The IAT most clearly measures automatic processing in the sense of capturing associations that are difficult to consciously control (see Teachman et al. 2012). To reflect a given association, such as a depression-related association that the self is associated with sadness, participants are presented with four categories that have been placed into two pairs on a computer screen: For instance, the category label Me is paired with Sad on the left side of the screen, and the category label Not Me is paired with Happy on the right side of the screen. Stimuli (words or pictures) from any of the four labeled categories appear one at a time in the center of the screen, and the participant is simply asked to indicate on which side of the screen the stimulus belongs. Thus, any word that fits in the category Me (such as I) or Sad (such as miserable) would be classified on the left side, whereas any word that fits in the category Not Me (such as them) or Happy (such as excited) would be classified on the right side. Participants are asked to do the classifications as quickly and as accurately as possible. After a series of classification trials, the category label pairings are switched, so Me is paired with Happy on the left side of the screen, and Not Me is paired with Sad on the right side of the screen. Participants are again asked to classify stimuli from any of the four categories as quickly and as accurately as possible. The difference in average reaction time for doing the classifications in the condition that matches depressive associations (when Me is paired with Sad, and Not Me is paired with Happy) versus the condition that contradicts depressive associations (when Me is paired with Happy, and Not Me is paired with Sad) indexes the strength of the implicit depression association. The order of the category pairing conditions is typically counterbalanced so that half of the participants complete the depression-congruent condition first and half complete the depression-incongruent condition first.

The task is indirect in that people are never explicitly asked whether they identify themselves as sad; this association is inferred from the difference in average reaction time across conditions. Further, the task reflects uncontrollable processing because the speed of categorizing stimuli is influenced by the category pairing, although that should not be relevant. For example, classifying the stimulus miserable in the Sad category should not be influenced by whether Sad is paired with Me (the depression-congruent condition) or with Not Me (the depression-incongruent condition), but the pairing clearly matters, such that people are faster to do the classifications when the categories are paired in a way that matches their automatic associations. This suggests that the classification speed is difficult to consciously control and that the paired categories can interfere with the classification. Thus, a depressed person is expected to be faster at categorizing stimuli during the depression-congruent condition than a nondepressed person is because that category...
pairing causes less interference for the depressed person given that she associates herself with being sad.

An advantage of the task is that it is easy to adapt to reflect a wide variety of implicit psychopathology associations by simply switching the category labels and associated stimuli (see Roefs et al. 2011, Teachman & Woody 2004, Teachman et al. 2010). Moreover, the task can be completed in just a few minutes and has been shown to reliably and incrementally predict a wide variety of outcomes across fields (see recent meta-analysis by Kurdi et al. 2018). The disadvantages of the task include the relative nature of the associations being assessed; in the depression example, the task must be interpreted as implicit sad versus happy associations with the self versus others. Thus, even if the research question only concerns the sad + self association, this association cannot be interpreted in isolation from the relative comparison categories (Happy and Not me). Further, the task seems better able to capture conceptually simple associations rather than sophisticated, complex ones [in line with Greenwald’s (1992) discussion of the dumb unconscious]. As an example, when psychopathology researchers have tried to measure complex clinical concepts with an IAT, such as the catastrophic misinterpretation of bodily sensations, which is central to anxiety sensitivity and cognitive models of panic disorder, null results are common (e.g., Steinman et al. 2011).

The task also sometimes has limited internal consistency, although it is reasonably reliable for a reaction-time task (see Gawronski & De Houwer 2014). Among other criticisms, there have also been questions about the extent to which the associations measured reflect awareness of cultural views versus individual associations (e.g., Arkes & Tetlock 2004) and concerns about the validity of the IAT’s scoring algorithm (e.g., Blanton et al. 2006) and its susceptibility to faking (e.g., Fiedler et al. 2006). While these criticisms must be considered carefully, it is worth noting that some of the critiques are more problematic in the context of an IAT measuring racial bias or other associations about marginalized groups than in the psychopathology context (e.g., concerns about awareness of cultural views or stereotypes are less relevant when measuring associations with the self and sadness or anxiety compared with measuring associations about race).

A number of variants of the IAT have also been introduced, including a variety of approaches that try to reduce the relative nature of the task, such as a Single Category IAT (e.g., Karpinski & Steinman 2006) and a Brief IAT (Sriram & Greenwald 2009). In the Brief IAT, stimuli from a relative comparison category are usually included as background stimuli (so the task still includes relative components), but no explicit comparative category label is included. For example, in a depression Brief IAT, the category label Happy may not be presented, even though participants would have to indicate whether happy-related stimuli, such as cheerful, should be categorized into the Sad category (when they are paired with Me versus with Not Me), and the task often uses fewer trials than a regular IAT. Another variant of the IAT, the GNAT (Nosek & Banaji 2001) is also a relative implicit association measure, but it uses a different classification approach than the IAT. On the GNAT, participants indicate whether a stimulus belongs (is a Go) or not (is a No-go), rather than classifying stimuli on the left or right side of the screen, and the GNAT often uses signal detection analysis of error rates with a brief response window. Also, a small number of studies in the psychopathology literature have used the IRAP (Barnes-Holmes et al. 2010), which deviates more substantively from the IAT because it assesses propositional beliefs rather than the simpler associations typically captured with the IAT and more closely related implicit association measures. [While it is not clear to what extent the IRAP captures the use of memory representations that do not require conscious introspection or control, we include it here given that it has been applied in interesting ways in the depression and obsessive–compulsive disorder (OCD) literatures.] Not surprisingly, there is less evidence for the psychometric properties of these alternative approaches, but they can be useful variants, depending on the research question (see the review in Teige-Mocigemba et al. 2010).
A different class of measures uses priming to assess implicit associations. In a typical Evaluative Priming task (also referred to as Affective Priming; Fazio et al. 1986), participants are presented with a prime stimulus (either above or below conscious awareness) prior to being presented with a target stimulus that is positive or negative (e.g., a photo of an alcoholic beverage is presented before the word excellent). Participants are asked to decide whether the target stimulus is positive or negative (in this example, excellent is positive). Even though they are not asked to evaluate the prime (picture of alcohol), the speed with which they can make that evaluative or affective judgment of the target stimulus is influenced by the extent to which they associate the prime stimulus with positivity or negativity. Continuing the alcoholic beverage example, if a participant has a positive association with alcohol, he will be faster to judge the word excellent as positive after seeing the alcohol prime than a person who has a negative association with alcohol. Nice features of this task are that it can be adapted to provide an evaluation of almost any object and that the prime can be presented above or below the threshold for conscious awareness. Also, it is possible to calculate a separate priming effect for a positive association and for a negative association (so it is less constrained by the relative comparison feature that characterizes the IAT). A serious disadvantage of the priming tasks is that they tend to have quite low reliability (Gawronski & De Houwer 2014). The reliability concern also applies to variants of Evaluative Priming, such as the Semantic Priming task (Wittenbrink et al. 1997), which presents a prime stimulus and then asks participants to do a lexical decision task that involves judging whether target stimuli (which are either real words or meaningless letter strings) are real words.

A number of different types of priming measures have been used in intriguing ways in the implicit social cognition field (e.g., the Affective Misattribution Procedure; Payne et al. 2005), but these measures have yet to be used extensively with clinical populations. Thus, as noted earlier, this brief discussion of measures does not exhaust all of the paradigms used to assess implicit associations, but it provides an overview of those measures used most widely to capture implicit psychopathology associations. This review also does not include tasks that mainly assess features of automaticity other than associations that are difficult to consciously control (e.g., we do not review attention bias measures that focus on the unconscious processing of disorder-relevant stimuli or dual-task paradigms that focus on efficient or capacity-free processing of disorder-relevant stimuli among other tasks). More generally, given that there are already many excellent, recent reviews of the attention bias literature in different clinical domains (e.g., alcohol: Christiansen et al. 2014; anxiety: McNally 2018; depression: LeMoult & Gotlib 2018), we focus on measures of implicit associations.

THEORETICAL AND EMPIRICAL ROLES FOR IMPPLICIT COGNITION IN PSYCHOPATHOLOGY

Cognitive theories of psychopathology from the past two decades regularly include a critical role for implicit cognitive processing. For each disorder, we briefly describe key examples of those theories and the resulting models and how they have evolved over time, and we also consider the evidence from implicit association measures in that disorder domain. The expectation is typically that disorder-relevant stimulus + evaluative outcome associations should be greater among symptomatic (versus control) participants, should predict disorder-relevant outcomes, and should be responsive to recovery following treatment. Across numerous studies, these hypotheses have been supported. For instance, a series of very large online studies (N > 10,000 participants for each study) demonstrated small-to-moderate relationships between self-reported symptoms of anxiety, alcohol use, depression, and eating disorders and the corresponding, disorder-specific associations assessed with an IAT (e.g., me + anxious, me + drinker, me + sad, high-fat food +
Another large online study demonstrated that relatively greater implicit associations of self + anxious and self + depressed on an IAT predicted a reduced likelihood of remission from, respectively, anxiety and depression (Glashouwer et al. 2011). However, as the reader will see, the literature tied to implicit association measures and psychopathology symptoms is complicated, and there are many inconsistent results across disorders and paradigms.

A number of valuable reviews have synthesized different aspects of the literature, often focusing on specific disorders (e.g., addiction and substance use: Wiers & Stacy 2013, Rooke et al. 2008; alcohol use disorder: Lindgren et al. 2018b; anxiety disorders and major depressive disorder: Teachman et al. 2012), and see Roefs and colleagues (2011), which reviewed research on implicit psychopathology associations more broadly. Here, we discuss key findings and themes from these reviews and augment prior reviews with more recent literature. We emphasize empirical research published within the past decade, although the studies reviewed build upon earlier pioneering work in the field (e.g., see Egloff & Schmukle 2002 for early work using the IAT within the context of state and trait anxiety). Given space constraints, we focus on adult psychopathology (in adults ages 18 and older) and evaluate three key classes of findings: (a) comparisons between groups with elevated symptoms versus control participants, (b) relationships between implicit association measures and other disorder-relevant variables, and (c) intervention effects (this classification extends the valuable work of Roefs et al. 2011). Rather than reify the diagnostic categories in theDiagnostic and Statistical Manual of Mental Disorders, 5th edition (Am. Psychiatr. Assoc. 2013), and in line with recommendations from the National Institute of Mental Health’s Research Domain Criteria (see Insel et al. 2010), we include studies with samples that report a range of symptom severity.

**Anxiety: Theoretical Proposals**

In general, the basic cognitive model of anxiety posits that individuals with heightened anxiety exhibit selective attention toward potential threat cues and increased likelihood of interpreting ambiguous information in a threatening way (Beck et al. 1985), resulting in threat-oriented cognitions that exacerbate beliefs about danger and vulnerability. Over the years, researchers have refined cognitive models of anxiety to more clearly describe how various forms of processing interact. When considering healthy threat processing, Mathews & Mackintosh (1998) proposed that if a cue has been previously learned (or biologically prepared) as dangerous, the cue is processed involuntarily, that is, without requiring conscious reflection or control. However, if the cue is novel, then more conscious, deliberate processes determine whether a stimulus is a threat. Individuals can use conscious strategies to counteract involuntary interpretations (e.g., engage in cognitive restructuring, a standard component of psychotherapy); however, for individuals with anxiety, relatively greater conscious cognitive effort is thought to be needed to counteract negative automatic thoughts. Thus, this older model posits a fairly traditional split between automatic and strategic cognition.

Beck & Clark (1997) proposed a three-stage information processing model. First, an individual with anxiety identifies a potential threat based on an initial rudimentary evaluation and then involuntarily appraises that stimulus as a threat. This second stage is rapid and inflexible and largely outside of conscious awareness and control. This activation causes individuals to narrow their thinking to focus on the activated negative thoughts in an effort to protect themselves from danger. In the third stage, individuals engage in full semantic processing to identify how they will cope with the threat. While still proposing a somewhat linear process, this model reflects a shift away from a strict implicit–explicit distinction and considers different stages of automatic filtering and interpretation. The dual implicit process model (March et al. 2018) proposes a similar threat-processing system, whereby bodily threat is first rapidly and preferentially processed, then
implicitly evaluated for valence, and later evaluated in a relatively more controlled way than during the first two stages.

There are many other influential theories and models that describe how anxiety influences cognitive processing. For instance, processing efficiency theory (Eysenck & Calvo 1992) suggests that anxious worrying reduces working memory capacity for concurrent tasks, so individuals with anxiety will be less efficient (another feature of automatic processing) at performing such tasks. In subsequent refinements to that theory, researchers proposed the attentional control theory (Eysenck et al. 2007), which posits that individuals with anxiety allocate more cognitive resources to attention processes that prioritize detecting threats in the environment, thereby reducing the cognitive resources available for more productive, goal-directed attention. This theoretical approach does not directly focus on implicit associations, but the posited interference caused by overactive, involuntary threat-detection mechanisms aligns well with our construal of implicit associations as representations in memory that link a stimulus and unintentionally activated evaluative outcome and that do not require conscious reflection to influence affect, cognition, or behavior. More recently, Peschard & Philippot (2016) extended this work to more fully integrate working memory into the approach.

This small sampling of the many cognitive models of anxiety highlights how dual-process perspectives are shifting toward less linear, more dynamic understandings of the links between different features of automatic and strategic processing: a shift that we hope will continue to better recognize how the many components of threat processing interact in complex and dynamic ways.

**Anxiety: Empirical Evidence**

We consider the evidence for implicit psychopathology associations separately across specific phobias and panic, social anxiety, and generalized anxiety disorders.

**Specific phobias.** Generally, individuals with specific phobias demonstrate relatively stronger implicit associations between their feared stimuli and fear-relevant evaluative outcomes than non-phobic individuals. In the earliest demonstration of this phenomenon, students with spider (versus snake) fear were quicker to evaluate spiders (versus snakes) as more threatening on an IAT across a range of evaluative dimensions (danger, disgusting, afraid, bad; Teachman et al. 2001). This finding and others like it showed that the fear evaluations were occurring without requiring conscious reflection or control. There have been numerous replications of this basic pattern, with most research occurring within the context of spider phobia (see the review in Roefs et al. 2011). Some research has also shown that implicit phobia-relevant cognition may predict meaningful outcomes, including self-reported fear and avoidance behavior (e.g., Ellwart et al. 2006, Teachman 2007) and automatic physiological responses (e.g., Van Bockstaele et al. 2011, Woud et al. 2011). In some cases (e.g., Teachman 2007), the implicit associations also showed incremental predictive validity after controlling for prediction by more deliberate measures, pointing to their unique predictive role.

Recently, researchers have begun to consider the contextual factors that may influence responses to implicit measures among individuals with specific phobias or fears. For example, the influence of implicit processes may be more impactful on fear-relevant behavior as the ability to control these processes diminishes (Effting et al. 2016). This is in line with double-dissociation hypotheses that expect implicit cognition measures will predict outcomes more strongly when conscious control of the outcome is difficult, such as when the outcome occurs spontaneously or resources are depleted (see Schnabel & Asendorpf 2010, although see Kurdi et al. 2018). Finally, implicit measures of phobia-relevant cognition may be sensitive to intervention effects...
(although see Huijding & de Jong 2007). For example, an exposure treatment (Teachman & Woody 2003) and a cognitive defusion intervention (Ritzert et al. 2015) led to reductions in implicit phobia-relevant cognition, assessed with, respectively, an IAT and an IRAP.

**Panic disorder.** Individuals with elevated panic symptoms demonstrate relatively stronger implicit, panic-relevant cognition across multiple measures (see the review in Roefs et al. 2011). For example, when responding to a Semantic Priming task, individuals with panic disorder (versus nonclinical control participants) reacted more quickly to panic-relevant catastrophes (e.g., heart attack) when they were primed with anxiety symptoms (e.g., “You notice an increasing palpitation”; Schneider & Schulte 2007). Certain implicit measures also predict panic symptoms and are sensitive to treatment effects. For example, one study found that the strength of catastrophic associations, assessed using the modified Semantic Priming task mentioned earlier (Schneider & Schulte 2007), incrementally predicted changes in anxiety sensitivity (Schneider & Schulte 2008). Other research using the IAT found that changes in self-panicked associations predicted changes in panic severity across 12 weeks of cognitive behavioral therapy (Teachman et al. 2008). In a secondary analysis of this study, researchers used the quadruple-process model (Conrey et al. 2005), which decomposes the parameters in the IAT (and related measures) to better isolate the different automatic and strategic processes that contribute to the measure’s outcome. In this reanalysis of within-group change across the panic control therapy, researchers demonstrated that it may be possible to change automatically activated associations in memory as opposed to simply overriding those associations with strategic processing (Clerkin et al. 2014a).

**Social anxiety disorder.** Theoretically, individuals with heightened (versus low) symptoms of social anxiety disorder (SAD) should demonstrate relatively more negative self-social associations; however, empirically, the literature is quite mixed (see the review in Roefs et al. 2011). Gilboa-Schechtman et al. (2017) recently argued that the inconsistent findings are partly due to the fact that there are different types of self-evaluations, which vary in importance for socially anxious individuals. For example, when a Single Category IAT was used, individuals with SAD (versus nonclinical control individuals) demonstrated relatively weaker implicit self-evaluations tied to social rank (e.g., self-dominant; Gilboa-Schechtman et al. 2017). In contrast, there was not a significant group difference on implicit self-evaluations tied to social affiliation (e.g., self-friendly). Other research suggests that comorbidity (Wong et al. 2014), sex (Glashouwer et al. 2013), presence of a social threat (Hiller et al. 2017), and even working memory capacity (Salemink et al. 2013) likely influence the strength and direction of effects. As just one example, Glashouwer et al. (2013) unexpectedly found that relatively weaker implicit self-esteem (me + positive associations on an IAT) was associated with greater SAD symptom severity only for men (and not for women).

Recent research that focused on mechanisms may also be illuminating. For example, one study found that following a social threat induction (versus a nonthreatening control condition), individuals with SAD (versus a nonclinical control group) had relatively lower implicit self-esteem (self + positive associations on an IAT; Hiller et al. 2017). Further, within the SAD group, relatively lower implicit self-esteem was associated with greater state anxiety and, importantly, mediated the impact of the social threat induction on state anxiety. Finally, research suggests that it is possible to change implicit social anxiety cognition using a variety of different interventions, including computerized cognitive bias modification tasks (CBM; Clerkin & Teachman 2010, Schnabel & Asendorpf 2015), group treatment (Gamer et al. 2008), and cognitive defusion (Kishita et al. 2014).

**Generalized anxiety disorder.** Less research has focused on implicit measures of cognition within generalized anxiety disorder (GAD); however, research suggests that implicit measures may
be useful for understanding the phenomenology and treatment of extreme worry. For instance, pointing to the pervasive nature of worry within GAD, one study demonstrated that individuals with GAD not only displayed relatively negative implicit associations toward clear worry-related targets (e.g., infidelity + agitated on a Single Category IAT), they also displayed similarly negative associations toward more neutral targets (e.g., neighbors; Reinecke et al. 2010). In contrast, healthy control participants displayed negative implicit associations only with the worry-related targets. Finally, a recent study that used emotion regulation therapy to treat GAD and major depression found that changes on an IAT were associated with a range of meaningful outcomes, including changes in emotional clarity, negative emotionality, and quality of life (Renna et al. 2017).

In summary, across the anxiety disorders, there is substantial evidence that implicit associations are greater among symptomatic (versus control) participants, predict disorder-relevant outcomes, and are responsive to recovery following treatment. Although the literature is more mixed for some anxiety disorders (e.g., SAD) than others (e.g., specific phobias), the breadth of relevant literature is also greater for SAD than for other areas, and the emphasis on moderators has been more pronounced, so there has been more opportunity to identify nuances and contextual effects in the implicit associations literature for SAD.

**Posttraumatic Stress Disorder: Theoretical Proposals**

For individuals who do not recover in the weeks or months following a trauma, posttraumatic stress disorder (PTSD) is marked by repeated, unwanted experiencing of the event; avoidance of reminders of the event; and hyperarousal. At a general level, prominent cognitive models of PTSD (e.g., Ehlers & Clark 2000) posit that PTSD occurs when an individual processes the trauma such that there is still an impending threat. For example, two soldiers returning home from war may both have experienced the same bomb explosion. One of the soldiers was able to process the trauma in a way that allowed her to contextualize the event (i.e., bombing happens during war), and she was able to recover once returning home. However, the other soldier ruminated on the bombing and possibility of his death and, now home, is scared of reminders of the event (loud noises and shouting). Over time, the second soldier continues to reexperience the bombing and tries to avoid all reminders. However, his avoidance attempts are likely to be unsuccessful, as this model suggests that when an individual is reminded of the trauma, the threat is activated and symptoms of PTSD occur. Memories of the trauma are easily accessible, unintentionally retrieved, and readily triggered involuntarily by stimuli linked temporally to the trauma. Thus, in this model, traumatic memories share multiple characteristics with implicit associations. This model also highlights a more iterative relationship between unintentionally retrieved memories and reflective processing, in that memories and appraisals are described as reciprocal and not necessarily linear.

Another theory of PTSD differentiates between two types of memories formed after a traumatic event: Verbally accessible memories are able to be consciously processed and edited over time, whereas situationally accessible memories are not subject to verbal processing, but are involuntarily activated by internal (e.g., thoughts) or external (e.g., sounds) stimuli (Brewin et al. 1996). These situationally accessible memories overlap (as in the earlier model), in part, with the concept of implicit associations. When situationally accessible memories are activated, conscious strategies process the information in an effort to make sense of the memories. These processes can interact iteratively; the individual either continues to link trauma cues to conscious processing and tries to avoid trauma cues in the environment (perpetuating PTSD) or consciously processes the involuntary memories in a healthier way (thereby reducing negative affect). More recent theories seek to integrate neurological findings to support varying forms of cognitive processing within individuals with PTSD (e.g., Lanius et al. 2017).
Posttraumatic Stress Disorder: Empirical Evidence

In comparison to healthy control participants, individuals with PTSD showed relatively greater implicit self-disgust associations on an IAT (Rüsch et al. 2011), and compared with traumatized individuals without PTSD and nontraumatized individuals, individuals with PTSD displayed relatively more implicit self-guilt associations on an IAT (Bockers et al. 2016). This suggests that implicit cognition related to disgust, guilt, and negativity may predict PTSD diagnostic status.

Implicit measures may also be useful for predicting correlates of PTSD. For example, one study found that a self-traumatization IAT predicted PTSD symptoms over and above the extent of trauma exposure and explicit cognitive measures (Lindgren et al. 2013b). In line with this finding, as assessed by an IAT, trauma frequency and severity may be associated with stronger implicit associations of violence + negativity (Bluemke et al. 2017), self + anxiety and self + depression (Johnson et al. 2011, van Harmelen et al. 2010), and self + betrayal (Delker & Freyd 2017). Surprisingly, for participants who experienced prior sexual abuse, relatively greater sex + aggression associations on an IAT were negatively associated with past trauma frequency and current effects of trauma (Reed et al. 2011). More in line with expectations, individuals with a history of childhood physical abuse (versus those without this history) have relatively greater self + anxiety associations on an IAT (Rüsch et al. 2011). These findings suggest that abuse in childhood may have important, albeit complicated, direct or moderating effects on PTSD-relevant measures of implicit cognition. Finally, to our knowledge, no study has directly evaluated whether implicit trauma-relevant cognition is sensitive to treatment effects among adults.

Obsessive–Compulsive Disorder: Theoretical Proposals

Cognitive models of OCD posit that unwanted, intrusive thoughts occur among all individuals; however, individuals with OCD misinterpret these thoughts or images as being personally meaningful (e.g., Rachman 1997), leading them to recur and cause distress. For instance, an unwanted image of driving off a bridge is interpreted as a sign that the person is dangerous and wants to die and is likely to actually swerve the car. Surprisingly, given that OCD is characterized by involuntary cognitions, there are few theories specifically incorporating implicit cognitive processing in OCD. In a recent review of implicit beliefs in OCD, Ouimet and colleagues (2018) concluded that the current (and limited) state of the literature suggests that implicit (defined as unconsciously and rapidly activated concepts in memory) and explicit (defined as controlled, conscious, and slow) beliefs differentially influence behavioral and cognitive outcomes in individuals with OCD and that the data are mixed regarding whether changes in implicit cognition will change symptoms (Ouimet et al. 2018). A few published reviews focus on the neurological findings tied to implicit cognition in OCD, with one implicating the basal ganglia as a critical region of the brain for forming motor and cognitive patterns (Graybiel & Rauch 2000). The authors hypothesize that individuals with OCD will have compromised parallel-processing capabilities (of implicit and explicit processing) given that the corticobasal–ganglia pathways are critical to parallel processing, and individuals with OCD have abnormal functioning in the corticobasal–ganglia pathways.

Obsessive–Compulsive Disorder: Empirical Evidence

Given the nature of OCD’s core psychopathology, the challenges of using implicit measures to evaluate complex psychological constructs are especially pronounced. For example, one study used a series of IATs to demonstrate that individuals diagnosed with OCD (versus individuals with body dysmorphic disorder, SAD, and a healthy control group) had relatively stronger obsessive thoughts
+ shame associations (Clerkin et al. 2014b). However, other research has not found the anticipated between-group effects on implicit measures of OC-relevant cognition (e.g., self + aggression: Cludius et al. 2017) or relationships between self-reported symptoms and implicit measures (e.g., self + immoral, self + dangerous, unwanted thoughts + important: Teachman & Clerkin 2007, Teachman et al. 2006). Interestingly, studies using implicit measures other than the IAT to evaluate OC-relevant cognition have been more successful at confirming hypotheses (e.g., Brief IAT: Green & Teachman 2013). Research with other non-IAT paradigms suggests that implicit OC-relevant cognition is likely associated with a range of OC-relevant variables, including scrupulosity (GNAT associations: Pirutinsky et al. 2015), excessive responsibility and overestimation of threat (IRAP: Nicholson et al. 2013), and contamination fear and avoidance behavior (IRAP: Nicholson et al. 2014).

To our knowledge, there are no relevant treatment studies focusing on implicit OC-relevant cognition. However, there are intriguing results stemming from experimental validity studies. For instance, in one study, undergraduate students were (a) told that their unwanted thoughts were significant and indicative of their current values, (b) told that their unwanted thoughts were meaningless, or (c) not given any information about their unwanted thoughts (Teachman et al. 2006). In comparison with the control conditions (b and c), participants who were told that unwanted thoughts were significant displayed relatively stronger implicit evaluations regarding the importance of unwanted thoughts (unwanted thoughts + importance associations on an IAT). Surprisingly, the manipulation did not influence performance on IATs tapping self + immoral or self + dangerous associations.

Major Depressive Disorder: Theoretical Proposals

Beck et al.’s (1979) early cognitive model of depression suggested that individuals with depression express dysfunctional, distorted, negative attitudes about themselves and the world, which arise from negative attention, interpretation, and memory biases. These cognitive biases arise from depressive, automatically activated schemas, conceptualized as representations stored in memory that reflect rigidly held core beliefs, such as “I am worthless.” Schemas are a somewhat fuzzy construct but overlap in certain ways with the (also fuzzy) idea of implicit associations. However, Beck’s early proposals that automatic schema fairly directly result in negatively biased explicit cognitions suggest a more linear automatic-causes-strategic path than do current, more dynamic conceptualizations of implicit cognition.

Beever’s (2005) model also aligns more with dual-process models (as opposed to more dynamic approaches) and suggests that a cognitive vulnerability to depression arises out of a bias toward uncontrollably associating the self with negative attributes. If these automatic biases violate more consciously controlled expectancies (e.g., individuals do not explicitly endorse thinking about themselves as bad), it is possible to override the involuntarily activated cognitions. However, if these associations do not violate expectancies or if the individual does not have the cognitive capability to engage in corrective processing, then that individual may be more likely to develop depression. A nice feature of this model is that it directly considers how implicit (defined as “quick and effortless information processing”; Beever 2005, p. 977) associations have a causal role in the development and maintenance of depression.

Major Depressive Disorder: Empirical Evidence

Contrary to theoretical expectations that measures that capture the uncontrollable processing of emotional stimuli will be negatively biased in major depressive disorder (MDD) (see the review in
Teachman et al. 2012), individuals with elevated depressive symptoms do not consistently demonstrate negative implicit associations with the self (e.g., see the review in Roefs et al. 2011). In fact, across several studies using several different implicit measures (the IAT, the Name Letter Preference Task, and the Extrinsic Affective Simon Task), evidence was “unequivocally indicative” of positive implicit self-esteem (e.g., self + worth) among individuals with depression, which was either similar to, or in one case greater than, that of nondepressed control participants (De Raedt et al. 2006, p. 1,025).

That said, other findings are more consistent with theoretical expectations for negative implicit cognitive biases. For example, compared with healthy control participants, individuals with current depression demonstrated lower implicit self-esteem on an IAT (e.g., self + positive), but individuals with remitted depression did not (Risch et al. 2010). Further, a post hoc analysis revealed that among patients with remitted depression, those with three or more prior depressive episodes demonstrated significantly lower implicit self-esteem compared with those with fewer episodes. Similarly, using data from the Netherlands Study of Depression and Anxiety, researchers found that during a 2-year period self + depressed implicit associations on an IAT correlated positively both with an individual’s number of prior depressive episodes at baseline and with the duration of depressive episodes (Elgersma et al. 2013). Also, in data from the Netherlands study, greater implicit associations of self + depressed were significantly related to a reduced chance of depression remission (Glashouwer et al. 2012).

Taken together, these findings suggest that depression-relevant implicit associations will be stronger when activated during prolonged and repeated depressive episodes, and they may impair one’s ability to recover (although see Krujit et al. 2013). Clearly the findings are mixed, however, and suggest there are important unresolved questions about when, for whom, and under which conditions depression-relevant implicit associations will be evident. Important contextual factors include comorbidity status (e.g., Valiente et al. 2011, Van Tuijl et al. 2016) and whether a mood induction was used prior to evaluating implicit measures (e.g., Hussey & Barnes-Holmes 2012, van Tuijl et al. 2018b; see the review in Roefs et al. 2011). Additionally, the valence and types of stimuli used are critical. For instance, some research suggests that individuals with MDD experience problematic inhibition of negative information (e.g., Dai et al. 2011), whereas other studies point to the difficulty that depressed individuals may have in processing positive information (LeMoult et al. 2012). Along these lines, only individuals with mild or moderate depression symptoms (and not those low in symptoms) showed a reduction in positive emotional responses measured on an IRAP following a sad mood induction (Hussey & Barnes-Holmes 2012). Another intriguing line of research focuses on discrepancies between implicit and explicit self-esteem (e.g., Jabben et al. 2014, Smeijers et al. 2017). For example, one study found that high implicit (e.g., me + valuable on an IAT) and low explicit self-esteem predicted heightened depressive symptoms, suicidal ideation, and loneliness, but that the reverse (low implicit, high explicit self-esteem) predicted only loneliness (Creemers et al. 2013).

In terms of intervention effects for depressed populations, the picture is similarly complicated. Some research suggests that implicit measures are sensitive to recovery following intervention (e.g., Leyman et al. 2011, Price et al. 2014), whereas other research does not (Adler et al. 2015). Again, moderators and contextual factors are likely to be important in understanding discrepant findings across studies. For instance, a recent study found that among individuals with elevated depressive symptoms, those high on trait mindfulness demonstrated significant improvements in self + negative implicit associations on an IAT following a brief mindful acceptance induction (Keng et al. 2016). However, individuals low on trait mindfulness actually demonstrated worse self + negative implicit associations following the induction.
In summary, measures of implicit cognition are useful for understanding the phenomenology of MDD. However, to interpret this complicated literature, it is critical to consider contextual and interacting factors, such as the level of consciously endorsed self-esteem and the course of depression over time. Broadly speaking, it appears that negative implicit associations about the self are likely to be more pronounced during (versus after) an active major depressive episode (e.g., Risch et al. 2010). Further, as the number of prior depressive episodes increases, so too does the strength of negative implicit associations toward the self (e.g., Elgersma et al. 2013), and there are hints that the presence of relatively more negative implicit associations prior to treatment might impair the ability to recover (e.g., see Glashouwer et al. 2012). However, these summary comments should be interpreted with considerable caution given the sparse and mixed state of the treatment literature.

**Alcohol Use Disorder: Theoretical Proposals**

General cognitive models of AUD (e.g., Tiffany & Conklin 2000) hypothesize that with repetition, drinking alcohol becomes automatized in that drinking behaviors become difficult to control, stimulus bound, and stereotyped. Multiple models of AUD propose an interaction between impulsive and controlled cognitive systems (e.g., Lindgren et al. 2018b, Stacy & Wiers 2010). For instance, in applying the reflective–impulsive model to addiction, Deutsch & Strack (2006) suggest that someone who starts with neutral associations with alcohol may then involuntarily pair alcohol with positive attributes as he drinks more frequently. Alcohol use may provide positive reinforcement (e.g., euphoria) and negative reinforcement (e.g., distraction from distressing thoughts or negative mood). Thus, over time alcohol becomes more strongly (and involuntarily) linked to positive feelings, even if reflective attitudes differ. Moreover, alcohol use may lead to decreased cognitive control and reflective processing (see Wiers et al. 2013). In related work, Stacy & Wiers (2010) propose connectionist models, whereby associations can provide the basis for unconscious recollection or introspection, and neurological models, whereby the brain adapts or becomes sensitized to addictive chemicals, creating attentional and approach biases among individuals with substance use disorders. Further, in their review of neurocircuitry and addiction models, Koob & Volkow (2009) propose that addiction is characterized by three stages: intoxication, withdrawal, and preoccupation/anticipation. Impulsivity plays a larger part in the earlier stages, whereas compulsivity has a bigger role in later stages. Notably, this proposal recognizes ongoing roles for different aspects of automatic and strategic processing across different stages.

These various models are notable in that they propose a clear role for increasingly positive implicit associations with alcohol that both motivate and reinforce drinking behaviors. Further, multiple models of addiction explicitly recognize the dynamic, iterative interplay of automatic and strategic processing over time. Perhaps this is because it is critical for any model of AUD to explain why drinking escalates when a person may consciously wish to stop, and all-or-nothing, implicit-then-explicit, dual-process models are not adequate to explain this seeming paradox.

**Alcohol Use Disorder: Empirical Evidence**

Similar to findings for other disorders reviewed here, research suggests that certain implicit measures are especially impactful in the context of AUDs. For instance, across a number of studies, having a relatively strong implicit drinking identity (e.g., drinker + me on an IAT) incrementally predicted drinking outcomes over and above explicit measures (e.g., Frings et al. 2016, Lindgren et al. 2013a). Importantly, some studies have shown that implicit drinking measures, particularly measures of drinking identity, longitudinally predict outcomes (Lindgren et al. 2016b, 2018a).
Multiple reviews (Roefs et al. 2011, Stacy & Wiers 2010) have also noted that implicit positive alcohol associations on an IAT, but not negative alcohol associations, predict actual alcohol consumption over and above explicit measures (although see Dickson et al. 2013, Vilenne & Quertemont 2015). This may be because negative alcohol associations reflect societal norms, whereas positive alcohol associations reflect more personal alcohol beliefs and tendencies (see Stacy & Wiers 2010).

Meanwhile, findings for other implicit measures of alcohol-relevant cognition are more mixed. For instance, some studies suggest that drinkers’ implicit associations between alcohol and approach motivational states on a GNAT (Kreusch et al. 2013) and alcohol and arousal/excitement on an IAT (e.g., Houben & Wiers 2009) are meaningfully tied to drinking outcomes. In contrast, other studies failed to find the anticipated implicit associations between alcohol + approach and alcohol + arousal on IATs (e.g., Vilenne & Quertemont 2015). These discrepancies may be due to differences in populations (e.g., heavy drinkers versus non-heavy drinkers) or the presence of advertising in visual stimuli (see Kreusch et al. 2013). Other potential moderators that may influence the relation between implicit drinking-related cognition and outcomes include age (Davies et al. 2017), sex (Lindgren et al. 2016a), executive functions (Lavigne et al. 2017), mood (Lindgren et al. 2018c), distractibility (Farris et al. 2010), and other substance use (e.g., Cohn et al. 2014). For example, positive alcohol associations on an IAT predicted drinking behavior among adults in college aged 18 and older, but not among adolescents aged 11–17 years (Davies et al. 2017).

Finally, there is evidence that problematic implicit alcohol cognition can be modified. For example, among unselected participants, watching alcohol advertising is associated with increased implicit positive associations with alcohol (e.g., Zerhouni et al. 2016). Unfortunately, there is not consistent evidence that interventions aimed at impacting drinking behavior by directly changing alcohol-relevant implicit cognition do so effectively (for reviews see Lindgren et al. 2018b, Roefs et al. 2011). For example, although a self-monitoring intervention benefited participants’ performance on an affective priming measure of implicit alcohol cognition at a marginally significant level, participants’ actual alcohol consumption did not change (Maas et al. 2013). Nevertheless, it is worth highlighting that some studies that use CBM procedures to alter implicit cognition have been effective. For example, implicit cognition and drinking outcomes were improved by a working memory task targeting executive function (Houben et al. 2011), a GNAT training inhibitory responses to alcohol (e.g., Houben et al. 2012), and an Approach–Avoidance Task training avoidance of alcohol stimuli (e.g., Wiers et al. 2011). This suggests that CBM may be one method that can be used to alter implicit cognition and other AUD symptoms, although the parameters influencing the efficacy of these interventions, such as the need for accompanying traditional alcohol interventions, are not currently well understood [for interesting recent commentaries on this topic, see volume 79 of the Journal of Studies on Alcohol and Drugs (Babor 2018)].

In summary, there is robust evidence from measures of implicit associations that alcohol-related associations, particularly implicit drinking identity and positive associations with alcohol, are related to concurrent and longitudinal AUD outcomes. In contrast, the data supporting the clinical utility of other measures of alcohol-related implicit cognition, such as alcohol + arousal associations, are more mixed. Finally, interventions focused on changing alcohol-relevant implicit cognition, such as CBM, have shown some promise, but they have inconsistently impacted meaningful outcomes.

**CONCLUSIONS AND CALL FOR FUTURE RESEARCH**

With the advent of more sophisticated paradigms to evaluate implicit disorder-relevant processing and new designs that allow for direct tests of the causal relationships between change in cognitive biases, such as implicit associations, and change in symptoms, there is much that we can now say...
about the importance of implicit associations to many different forms of psychopathology. The evidence for these associations provides clues to help explain why some individuals see a harmless baby ribbon snake and barely give it a fleeting thought, while others get caught in a vicious cycle of increasingly elaborate threat processing that leads to a full-blown panic attack and the abrupt end to all camping trips. At the same time, in many respects, research on the theoretical and clinical implications of implicit psychopathology associations is in its infancy. We consider just a small subset of the many barriers that need to be overcome to make further progress in this field, and we end with a call for future research.

1. Implicit association paradigms need to advance so that measures have stronger reliability (both test–retest and internal consistency) and can more clearly assess associations that are less relative (e.g., assess associations about being sad that are distinct from associations about being happy). A related challenge is to create paradigms that reliably capture more complex associations; when introducing the IAT measure, we noted difficulties with assessing complicated ideas, such as the catastrophic misinterpretation of bodily sensations, and this challenge has arisen across many disorders.

2. Paradigmatic and quantitative analytic advances are needed to better isolate the various specific processes that are embedded within measures of implicit associations. For instance, the quadruple-process model (referred to as the quad model; see Conrey et al. 2005, Sherman et al. 2008) has been applied in the social cognition field to disentangle the automatic activation of associations aspect of implicit association measures, such as the IAT, versus the more controlled detection of correct and incorrect responses, the strategic override of bias, and guessing aspects of the measured associations [see also related work using the diffusion model (Klauer et al. 2007) and the process dissociation model (Payne 2008)]. However, these methods have rarely been applied to the psychopathology field (although see Clerkin et al. 2014a as one recent example). These approaches could be a natural fit in the clinical field given critical questions about which aspects of implicit associations need to change with treatment to reduce an individual’s vulnerability to relapse. For instance, this knowledge could be used to help clinicians decide when to end treatment to optimize the chances for their clients’ maintenance of treatment gains.

3. Paradigms need to capture temporal dynamics more effectively. Measures such as the IAT capture a static snapshot of associations at one point in time, but we really want to know the time course of the interactions among various automatic and strategic processing biases. Tasks such as the MouseTracker paradigm (see Hehman et al. 2015) and others that use continuous tracking [e.g., the Tracking Affect Ratings Over Time, or TAROT, task (Fua & Teachman 2017)] provide some initial steps for measuring more iterative processing. Applying these insights to assessing implicit psychopathology associations will hopefully open up new opportunities to redirect the seemingly uncontrollable cycles of escalating disorder-relevant processing.

4. More investigations are needed into changes in implicit psychopathology associations across time and following different types of treatment. Analogously, more evaluations are needed of the causal links between changes in implicit associations (e.g., through conditioning paradigms or other cognitive bias–modification approaches) and changes in symptoms and other aspects of functioning. Also, little is known about the contribution of implicit psychopathology associations to the etiology of clinical problems (although see important work predicting onset from the Netherlands Study of Depression and Anxiety by Glashouwer et al. 2011 and Kruijt et al. 2013, among others).
5. Tracking implicit associations over time also needs to be done in ways that have more external validity. This includes increasing the ecological validity of tasks (to move beyond classifying words on a computer screen) and integrating implicit association measurement into daily life (e.g., using ecological momentary assessments). This would allow for a clearer evaluation of the ways in which implicit associations interact with a nearly infinite list of potentially important state variables (e.g., current affect; situational demands; cognitive capacity and other aspects of executive functioning; intoxication level; social context, such as being alone or not; motivation level; the egosyntonic versus egodystonic nature of the associations; see Lindgren et al. 2018b, among many other examples).

6. Beyond methodological advances, it will be important for theoretical advances to occur as well. Although this is changing, traditional clinical science models are sometimes too narrow (e.g., considering only one feature of implicit cognitive processing as though that feature occurs in isolation) or too broad (e.g., using loose terminology that treats automatic/implicit processing as a single, all-or-none process, thus losing sight of the important differences between unconscious versus uncontrollable versus unintentional processing). Along these lines, just as measurement advances are needed that capture more than just a linear or static snapshot of a process that is actually highly iterative and interactive, so too are theoretical advances needed to apply and extend contemporary social cognition and neural models of dynamic processing to explain different forms of psychopathology.

7. Traditional models that consider the role of automatic processing in psychopathology are often disorder specific, omitting the important question of implicit associations that are transdiagnostic risk markers. For instance, it is easy to imagine how an association between the self and the concept inadequate could confer vulnerability to both depressive and anxiety disorders and set people up to believe they need to self-medicate with alcohol to meet some unrealistic, perceived performance goal. The consideration of transdiagnostic implicit psychopathology associations also aligns well with the National Institute of Mental Health’s focus on Research Domain Criteria (see Insel et al. 2010), which seeks to define transdiagnostic, dimensional markers of vulnerability to psychopathology.

8. More research is also needed to be able to place a given implicit association score in context. For instance, we know little about how individual differences in demographic characteristics and different backgrounds and identities (e.g., race, ethnicity, age, gender, education, nationality, treatment history) relate to the expression of implicit psychopathology associations (although see Werntz et al. 2016 for some suggestive findings). Analogously, it will be important to better integrate models of psychopathology with models of healthy emotion reactivity and regulation [e.g., Gross’s (1998) seminal process model of emotion regulation and recent considerations of implicit emotion regulation (Braunstein et al. 2017)]. This will enhance understanding of how implicit associations change on a continuum from healthy to pathological.

This future research wish list is clearly ambitious and will take time to fulfill, but the knowledge that may be gained is considerable. We already have hints that implicit associations in the clinical field can predict treatment outcomes (Teachman et al. 2008), likelihood of relapse (see van Tuyl et al. 2018a), and even risk of suicide (see Nock et al. 2010 and a recent review by Cha et al. 2018), among other important clinical outcomes. These hints suggest that investing in this research has strong potential to directly improve the lives of the millions of people suffering from mental illness.
SUMMARY POINTS

1. Many forms of psychopathology are characterized by implicit cognitive processing in which disorder-relevant information is processed involuntarily.

2. Implicit mental health associations (i.e., representations in memory that link a stimulus and involuntarily activated disorder-relevant outcome) do not require conscious reflection to influence affect, cognition, and behavior.

3. The evidence for implicit associations is more robust in anxiety and alcohol use disorders than it is in depression, and little is known about the causal role of implicit associations in disorder etiology and recovery.

4. Traditional models of psychopathology tend to reflect dual-process perspectives, which assume a relatively static automatic/implicit versus strategic/explicit split.

5. Newer psychopathology models are emerging and more updates are needed to better reflect dynamical systems perspectives in which mental representations are iteratively reprocessed, evolving continuously.

6. The methodological limitations of current measures of implicit associations have slowed theoretical advances because of difficulties in reliably measuring and isolating complex constructs and dynamic processes in ecologically valid ways.

DISCLOSURE STATEMENT

B.A.T. directs Project Implicit Mental Health and has a significant financial interest in Project Implicit, Inc., which provided services in support of some of the research cited in this review. The authors are not aware of any other affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors are thankful for the feedback provided by members of the Program for Anxiety, Cognition, and Treatment (PACT) lab at the University of Virginia. This research was supported by National Institute of Mental Health grant R01MH113752 grant to B.A.T.

LITERATURE CITED


Babor TF, ed. 2018. J. Stud. Alcohol Drugs 79(3)

Barnes-Holmes D, Barnes-Holmes Y, Stewart I, Boles S. 2010. A sketch of the Implicit Relational Assessment Procedure (IRAP) and the relational elaboration and coherence (REC) model. Psychol. Rec. 60(3):527–42


Dickson JM, Gately C, Field M. 2013. Alcohol dependent patients have weak negative rather than strong positive implicit alcohol associations. *Psychopharmacology* 228(4):603–10


9.24 Teachman et al.


Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. 2011. Retraining automatic action tendencies changes alcoholic patients’ approach bias for alcohol and improves treatment outcome. *Psychol. Sci.* 22(4):490–97


