The when and how of evaluative readiness: A social cognitive neuroscience perspective

Melissa J. Ferguson Michael T. Wojnowicz

Cornell University University of Washington

**Abstract**

In this article we consider the phenomenon of evaluative readiness, whereby the activation in memory of a goal leads to an unintentional increase in positivity toward stimuli that can facilitate the goal. We review four lines of work that together address the question of *when* goals lead to this kind of automatic shift in people's attitudes. We then consider *how* contemporary models of cognition might explain this effect. We review whether dual systems models and single interacting system models can explain the phenomenon of evaluative readiness. Based on recent work in cognitive psychology and computational neuroscience, we then argue for the potential explanatory value of turning to a multiple interacting systems framework for explaining the phenomenon of evaluative readiness.

**Introduction**

How does the activation of a goal in memory change how we interpret the world around us? Most research on goal pursuit and motivation has traditionally focused on how a goal can lead to changes in our *explicit* thoughts, plans, and emotions (e.g., Atkinson, 1964; Ajzen, 1991; Bandura, 1986; Carver & Scheier, 1998; Locke & Latham, 1990; Mischel, Cantor, & Feldman, 1996). This literature shows that once we enter into a (conscious) goal state (e.g., wanting to compete, be social, relax, achieve at work, etc.), we experience a different world – one that is more versus less aligned with that goal. We make intentions, plans, experience emotions, and behave differently.

And yet, recent research has revealed that the activation of a goal also triggers a wide range of changes that are less noticeable, less intentional, and less reportable (i.e., implicit). These include implicit effects on knowledge accessibility, judgment, perception, and attitudes (e.g., for reviews see Bargh, 2007; Fishbach & Ferguson, 2007; Förster & Liberman, 2007; Shah & Gardner, 2007), and even though such changes may operate under a person's conscious radar, and proceed unintentionally, they can nevertheless influence and predict how the person responds behaviorally to the environment (e.g., Balcetis & Dunning, 2006; Ferguson, 2007, 2008; Fishbach et al., 2003; Shah, 2005; Shah & Kruglanski, 2003). This line of work demonstrates that the activation of motivational states leads to undetectable and unintentional shifts in how we see, feel, and act in the world.

In this paper, we consider research on the effects of goals on people's implicit attitudes. Recent work shows that the activation of a goal in memory automatically shifts people's attitudes toward activities, events, and objects in ways that facilitate the pursuit of the goal (Ferguson, 2008; Ferguson & Bargh, 2004; Fishbach & Shah, 2006; Fishbach, Zhang, & Trope, 2010;

Moore, Ferguson, & Chartrand, 2011; Seibt, Häfner, & Deutsch, 2007; Sherman, Presson, Chassin, Rose, & Koch, 2003). This work shows that as soon as a goal becomes activated in memory, whether consciously or nonconsciously, people can automatically become "evaluatively ready" to pursue it.

Given the predictive validity of implicit attitudes for subtle and overt behavior (for reviews see Petty, Fazio, & Brinol, 2007; Wittenbrink & Schwarz, 2007), the finding that a goal automatically increases people's liking for stimuli that can help them reach the goal is a promising explanation for how the activation of a goal translates into goal behavior. In order for the activation of a goal to translate into goal-relevant behaviors, it seems necessary for the person's likes and dislikes to shift in order to be more in line with the goal. Ideally, people should have greater liking for those things that can move them toward the goal, and greater disliking toward those things that might distract them.

After reviewing the evidence for evaluative readiness in the first section of this paper, we then move on in the second section to a consideration of a model of cognition that might be able to explain such an effect. We briefly review the major tenets of standard dual-process perspectives of cognition currently popular in social cognition, as well as dynamical systems frameworks that are largely popular in cognitive sciences but not in social psychology. We describe how the phenomenon of evaluative readiness poses problems for each type of model, and conclude by speculating on how moving towards an integrated, neurobiologically-plausible model of cognition might best handle evaluative readiness.

## Section 1: *When* goals shift evaluations

In one sense, the claim that a person's current goal influences what that person most likes in the environment tends toward the circular. If a goal is defined as a *desired* end-point that fluctuates in its accessibility in memory (see Ferguson & Cone, in press), then the activation of a goal (e.g., academic achievement) should, at the very least, increase the person's positive regard for the goal itself (the concept of academic achievement). In other words, if a goal is broadly defined as a diverse array of knowledge related to the means and strategies for pursuing and meeting a *desired* end-point, then any increase in a person's positive regard for the end-point or even the means closely associated with that end-point could easily be construed as evidence that the goal has been activated, rather than that the goal has influenced some other presumably independent construct such as attitudes. Early research showed that when people were hungry they rated sucrose samples as more positive compared with when they were not hungry (Cabanac, 1971). In some ways, according to contemporary social-cognitive definitions of goals (see Bargh, 1990, 1997; Fishbach & Ferguson, 2007; Kruglanski, 2002), this evidence by Carbanac (1971) might be primarily interpreted as indicating that the participants were, in fact, hungry.

In another sense however, the claim that goals can influence what people like and dislike can be conceptualized and tested in ways that do not tend toward the tautological. In this first section of the paper, we identify four ways in which research has examined the basic claim that goals influence people's attitudes. This overview is meant to provide a brief sketch of the evidence for this phenomenon.

**I. Can goals shift how we implicitly evaluate stimuli?**

The field of social cognition has assumed that attitudes can be measured in two different ways – explicitly (i.e., directly) versus implicitly (i.e., indirectly). Explicitly measured attitudes are those that a person deliberately (knowingly) reports, usually on a Likert type of scale. Implicitly measured attitudes are those that are captured using covert paradigms in which the person does not realize that her or his attitudes are being measured (e.g., see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Ferguson & Fukukura, in press; Greenwald & Banaji, 1995; Petty et al., 2007; Wittenbrink & Schwarz, 2007). Researchers have lately argued that explicitly and implicitly measured attitudes are related but distinct constructs (e.g., Cunningham, Preacher, & Banaji, 2001; Greenwald & Farnham, 2000; Nosek & Smyth, 2005), even when accounting for method variance. Importantly, researchers at first adopted different assumptions about the context-dependence of explicit versus implicit attitudes in that the former was assumed to be more sensitive to various contextual factors than the latter (e.g., see Bargh et al., 1992; Bargh et al., 1996; Fazio et al., 1995). Implicit attitudes were assumed to be relatively impervious to variations in the context in which a perceiver encounters the stimuli that are evaluated. Given these assumptions, whereas goals would be expected to influence people's explicit attitudes, they would not be expected to influence people's implicit attitudes. From this perspective then, some of the initial research showing the effect of goals on *implicit* attitudes directly addressed (and challenged) these assumptions of the context-independence of implicit attitudes.

In one line of research, Sherman et al. (2003) asked smokers to either take a smoking break or refrain from smoking. Participants then completed a number of measures that captured their implicit attitudes toward smoking paraphernalia (e.g., cigarettes). The results showed that among chronic, heavy smokers, those who had *not* just smoked exhibited more positive attitudes than those who had just smoked. These results show that people who had yet to meet their need (in this case, to smoke) exhibited significantly more positive attitudes toward stimuli related to that need compared with those who had just successfully (and presumably enjoyably) satisfied that need.

In another line of work, Ferguson and Bargh (2004) asked participants to play a novel word creation game and either gave them a goal to do well on the game or did not. Participants then played the game for a few minutes, and then completed an ostensibly unrelated computer task that surreptitiously measured their implicit attitudes toward stimuli relevant to high performance in the game (e.g., *points, words, creative*). As they started the implicit attitude measure, half of the participants believed they had finished the word game, and the other half believed that they would be playing another round of the game after the computer task. The results showed that only those who had the goal to do well on the game *and* who believed they were still actively involved in the game displayed significantly positive implicit attitudes toward the game relevant stimuli. Other research has found similar types of effects of currently active goals on people's implicit attitudes toward stimuli related to the goal (Seibt et al., 2007).

Based on a burgeoning literature showing the various contextual factors that can alter the way in which people implicitly evaluate the stimuli around them (for a review see Wittenbrink & Schwarz, 2007), researchers have now largely agreed that implicit attitudes are highly sensitive to the context in general, including a person's temporary and chronic goals. Thus, the fact that an active goal can alter people's implicit attitudes speaks to the nature of implicit attitude generation (see Ferguson & Bargh, 2008; Ferguson & Porter, 2009). Also, however, such findings suggest that the activation of a goal has pervasive effects –influencing even our affective reactions that are generated within milliseconds of encountering stimuli. In this way, even our attitudes that are generated within milliseconds of encountering the corresponding stimuli are goal-dependent.

## II. Are goals more likely to influence explicit versus implicit evaluations?

Some of the research on the effect of goals on attitudes has focused exclusively on explicitly reported attitudes, such as the aforementioned work by Cabanac (1971). Fishbach and colleagues (Fishbach, Shah, & Kruglanski, 2004) have also shown that when people are actively pursuing an end-state, the positive affect they associate with the end-state can sometimes transfer to the means to reach the end-state. And, recent research by Fitzsimons and colleagues (Fitzsimons & Fishbach, 2010; Fitzsimons & Shah, 2008) shows that people even increase their positive, explicit evaluations of their relationship partners depending on whether those partners are instrumental for a currently accessible goal. This work together shows that when people are in an active goal state, they at times self-report more positivity toward those stimuli that help them to reach the goal.

And yet, interestingly, there is also evidence showing that goals sometimes do not influence both explicit and implicit attitudes. In several papers, findings indicate that active, conscious goals *do not* change people's explicitly reported positivity toward goal-relevant stimuli, but *do* change people's implicit attitudes toward the same stimuli. Sherman et al. (2003) found that even though heavy smokers who still needed their "fix" showed relatively more positive implicit attitudes toward smoking-related stimuli, they did *not* show any change on their explicit attitudes, compared with those who had just smoked. Ferguson and Bargh (2004) also found that participants who were playing the word game and wanted to do well, and also thought that they would be playing another round of the game, exhibited more positive implicit attitudes toward stimuli related to the game, but did *not* change their explicit attitudes toward the same stimuli. Similarly, Ferguson (2008) found that those with a conscious goal to achieve academically did not alter their explicit attitudes toward instrumental means to reach the goal even though they showed significantly more positive implicit attitudes toward those same stimuli.

This is an interesting dissociation because why would those with a conscious or accessible goal to reach an end-state not report greater positivity toward those stimuli highly associated with that end-state? This is especially intriguing because none of the goals in these cases seemed obviously controversial in any way, and it would seem as though there would not be any social desirability pressures to report a certain amount of positivity toward stimuli such as *school* or *game* or *letters*. So, why the dissociation? One possibility is that people simply have a hard time introspecting accurately on how positively they feel about stimuli at any given moment (e.g., Nisbett & Wilson, 1977; Wilson, 2002; Wilson & Dunn, 2004). That is, they may be able to say easily and with great accuracy that they feel positively versus negatively in general toward a certain end-point, but they may not be able to precisely intentionally identify *how* positively they feel. Thus, the degree of positivity people report may not always map on to the amount of positivity that has actually been implicitly activated in memory by the goal state in that moment. The amount of positivity that is unintentionally and rapidly activated when they first encounter a stimulus in reality or as depicted in a photo or verbal description, on the other hand, may at times map more precisely onto the fluctuation of positivity over seconds or minutes as the importance of goals waxes and wanes. This suggests that even though there may be times when a goal alters people's overt and explicit attitudes (e.g., Fitzsimons & Shah, 2008), goals may be more likely in some cases to reliably influence people's spontaneous and rapidly generated attitudes.

## III. Can non-conscious goals shift our evaluations of non-conscious stimuli?

Traditionally, a major assumption in the goals literature has been that people pursue goals consciously, with deliberation and intention about reaching the desired end-state, with various forms of conscious monitoring of their progress along the way (e.g., Atkinson, 1964; Ajzen, 1991; Bandura, 1986; Carver & Scheier, 1998; Locke & Latham, 1990; Mischel, Cantor, & Feldman, 1996). However, more recently researchers have argued and shown that people can pursue goals nonconsciously and unintentionally (e.g., Aarts & Dijksterhuis, 2003; Aarts, Hassin, & Gollwitzer, 2004; Bargh, 1990; Bargh et al., 2001; Chartrand & Bargh, 1996; Ferguson & Cone, in press). For example, people who were subtly primed with words related to achievement then displayed behaviors that met the classical criteria for motivational behavior, including persistence, resumption after an interruption, and an increase in the strength of the goal over time until it had been met (Bargh et al., 2001). Given this recent evidence, a question is raised as to whether even nonconscious goals influence people's likes and dislikes for the stimuli around them. In other words, what are the processing requirements for the activation of a goal to influence people's attitudes? Does a person have to be intentionally pursuing a goal in order for that goal to change what that person most likes in her or his environment?

Recent research (Ferguson, 2008; Moore, Ferguson, & Chartrand, 2011) suggests that the answer is no. In one study, participants who were subliminally or subtly exposed to stimuli related to a goal (e.g., academic achievement) showed more positive implicit attitudes toward stimuli relevant to the goal (e.g., *library* and *books;* see Ferguson, 2008). Thus, the mere activation of a goal, whether conscious or nonconscious, is enough to shift people's evaluations of the stimuli around them in a goal-consistent fashion. It is also worth noting that this work again only showed effects on people's implicit attitudes, and not their explicit attitudes.

But what about the stimuli that are being evaluated? Even if the goal itself has been activated outside of consciousness, perhaps there needs to be at least some conscious processing of the stimuli in order for people's evaluation of them to shift according to the goal. Previous research on the effect of goals on attitudes used implicit measures that ensured that the attitudes toward the stimuli were activated spontaneously and rapidly. However, they were also consciously processed in these measures, and therefore, there is the possibility that participants developed some strategic processing style that allowed them to realign their attitudes toward the stimuli according to their recently activated goal (see Klauer & Mierke, 2005; Klauer, Schmitz, Teige-Mocigemba & Voss, 2010; Mierke & Klauer, 2001; Payne, 2005). In the studies by Ferguson (2008), though, participants were not even consciously aware of the stimuli themselves and yet their current goal influenced their evaluations of those stimuli. This means that people may have active goals of which they are not aware which then influence their attitudes toward stimuli they do not consciously notice.

## **IV. Does *any* goal alter *any* person's implicit evaluations?**

Importantly, what are the moderators that constrain and define the nature of the relationship between goals and attitudes? Is it the case that the activation of any particular goal will influence any particular person's implicit attitudes? The research by Ferguson (2008) also provides initial evidence that only people who have some moderate amount of success at a given goal will show an increase in the positivity of their attitudes toward goal-relevant stimuli once the goal has been activated (and, see also Fishbach et al., 2003). For example, those participants with higher versus lower GPAs showed a change in the positivity of their implicit attitudes toward school-related stimuli after being primed (consciously or nonconsciously) with the academic achievement goal. Those who had a low GPA did not show this evaluative readiness effect. This indicates that when goals are easy to meet, such as sating thirst, playing a simple word game, or smoking a cigarette (for smokers), then most people on average will show a goal-consistent shift in their implicit attitudes once the goal has been activated. However, when the goal is more difficult, and when performance at it is more variable across people, then it seems to be the case that only "goal experts" show this kind of evaluative readiness shift.

Goal expertise is an interesting moderator because it suggests (based on correlational data) that perhaps this kind of an effect of goals on implicit attitudes is functional for the person. In other words, those people who immediately "see" the world more in line with their currently active goal, at least in evaluative terms, should be more likely to meet the goal. It makes sense that implicitly becoming "evaluatively ready" to pursue a goal once it has been activated could actually facilitate the pursuit of the goal. After all, those who show a greater immediate and spontaneous liking for activities, events, and objects that could help them reach the activated goal should be more likely to actually approach those things and pursue the goal. This kind of goal-induced realignment of what one most likes or dislikes in the environment could provide a performance edge in actually meeting the goal. This idea remains an open empirical question.

It is important to note that in this recent line of research, the differences (in their implicit attitudes) between those participants who were classified as experts at the goal and those who were not experts was *not* explained by differences in their reported motivation to reach the goal, the perceived instrumentality or relevance of the stimuli being evaluated, or their mood.

**Conclusion of Section 1**

These four lines of work provide a summary of some ways in which the claim that goals influence attitudes can be tested in ways that are not merely circular. We now move to a discussion of *how* the activation of a goal may influence a person's implicit (or explicit) attitudes. This question is just starting to be addressed and so we speculate here on how a specific model of cognition might apply to this particular empirical phenomenon. In particular, we consider recent work in cognitive psychology and computational neuroscience.

## Section 2: *How* goals influence evaluations

Evaluative readiness suggests a great deal of fluidity between flexible strategic goals and implicit attitudes. For instance, when an undergraduate student decides to head for the library to study for a final exam, her temporarily adopted strategic goal would cause her to experience greater implicit positivity towards libraries, books, and parties; however, just minutes later, if the student decides to step outside for a smoke break to make new friends, her newly adopted strategic goal would cause her to experience greater implicit positivity towards cigarettes. Current theoretical models of real-time mental processing do not readily explain this "strategic fluidity" -- the transient nature of strategic goals and their influence on implicit attitudes. In this section, we discuss how dual systems frameworks (from social psychology) and single interactive system frameworks (from cognitive science) both face problems in explaining evaluative readiness. We then describe a recently developed hybridized framework -- multiple interacting systems -- and sketch a description for how evaluative readiness could be explained from within this framework.

### I. Dual Systems: Strict Separation of Strategic Goals and Implicit Attitudes

Social psychological theorizing frequently refers to a distinction between two distinct computational systems with qualitatively distinct operating principles (Deutsch & Strack, 2000; Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006; Smith & Decoster, 2000): one system is implicit, associative, uncontrolled, fast, subconscious, and unintended (“System I”; Kahneman & Frederick 2002), and the second system is explicit, rule-based, controlled, slow, conscious, and unintended (“System II”).

According to the dualistic perspective, implicit attitudes and strategic goals are thought to emerge from the associative system and the rule-based system, respectively. Dual systems theory would posit that flexible strategic goals require sophisticated computational machinery and thus must rely on the rule-based system (e.g., Strack & Deutsch 2004). Goals have been theorized to consist of *future* oriented plans or behaviors to reach desired end states (e.g., Strack & Deutsch 2004), and such a construct could not live in the associative system. Associations form between two concepts for two reasons: either the concepts have "structural similarity" (i.e. they are determined share many semantic features) or they have "temporal contiguity" (i.e. they frequently co-occur in time). So the associative system has been assumed to be a slowly evolving experiential store, tracking a continuously updated set of links between concepts. As a result, it is said be inherently *reproductive,* with its computations simply reproducing past experience (e.g., Smith & DeCoster, 2000; Smith, 2007; cf. Amodio & Ratner, 2011). The conclusion from dual systems models has been that the associative system is fundamentally incapable of representing the future, and therefore inadequate for representing future goals (see Sloman, 1996; Strack & Deutsch, 2004).

In contrast, the rule-based system is inherently the right place for strategic goals to live. Unlike the primitive associative system, the rule-based system has the property of "syntactic binding." That is, the rule-based system can bind its concepts into thematic roles (for example, the logical roles found within the logical expression "x and y, therefore z"; or the syntactic roles found a linguistic expression with an agent, verb, and patient). As a result, the rule-based system has systematic structure: whereas the associative system could do nothing more than link the words "Mary", "John," and "love" without any sense of directionality, the rule-based system can bind these words into sentences with very different meanings, "Mary loves John" vs "John loves Mary" (Fodor & Pylyshyn 1988). Using syntactic binding, the rule-based system can transcend its experiences -- it can perform logical syllogisms, reason through higher-level mathematics, and achieve equal accuracy with familiar vs. unfamiliar material (Smith, Langston & Nisbett, 1992). Thus, the rule-based system is the only system that is *productive* (i.e. it can produce new thinking rooted in the future rather than the past). As a result, it has been concluded that the rule-based system must be setting and enforcing strategic goals. Only the rule-based system allows people to symbolically reason about future events that have never occurred, rather than acting reflexively in accordance with their experiential store (Strack & Deutsch 2004; Sloman 1996).

On the other hand, implicit attitudes are assumed to live in the associative system. In fact, by definition, implicit attitudes are assumed to share the properties of the associative system. Their defining characteristics include that they form by environmental conditioning (e.g., De Houwer et al., 2009; Olson & Fazio 2001), that they are more emotional than logical (e.g., Epstein, 1999; Rudman 2004), and that they summarize the valence of automatically activated knowledge structures (e.g., Gawronski & Bodenhausen 2001). Note that these defining properties are very un-goal-like. Because implicit attitudes are assumed to form via environmental conditioning, they are determined by uncontrollable transitions within the environment, and they do not necessarily reflect internal strategic goals. Because implicit attitudes are assumed to be more emotional than logical, they serve animalistic drives, short-term temptations, and momentary distractions much better than the rational calculus inherent in some long-term goal pursuits. Because implicit attitudes reflect a summary of the biases that automatically pop out during a knowledge activation stage, they describe a person's preferences *before* the arrival of a separate response selection mechanism that takes a person's strategic goals into account.

So strategic goals and implicit attitudes are separated by dual systems -- these two psychological constructs live in different systems with very different computational structures. The distinction is often applied to self-control research, which often emphasizes moments when the dual systems *conflict*. Social psychological theories of self-control postulate that people sometimes fail to meet their goals because they have trouble subordinating their wild, unruly associative system (filled with its temptations, emotions, cognitive heuristics, mere associations, and momentary distractions). On the other hand, evaluative readiness emphasizes moments when the dual systems *successfully communicate.* Note that in the evaluative readiness studies, a deliberately adopted strategic goal -- such as the desire to win a game or to perform well academically -- can moderate implicit attitudes within a matter of milliseconds. So evaluative readiness poses an especially interesting question of interactions between systems. That is, if implicit attitudes live in system 1, and strategic goals live in system 2, how do these systems communicate so seemingly flawlessly (or at all)? Despite many experimental findings of cross-talk between systems (e.g. Zajonc 1980; Blair, Ma & Lenton 2001), it is not currently understood how rules and associations would interact (e.g., see Greenwald & Nosek 2008). Because of this impasse, it is hard to for dual systems theory to help explain how evaluative readiness (among countless other examples of psychological phenomena) works.

## II. Single Interactive System Models: Interactive Dynamics Without Any Goals

The term "single interactive system" is meant to encompass an influential class of neural network models such as Normalized Recurrence (NR; Spivey 2007), Simple Recurrent Network (SRN; Elman 1990), Dynamic Field Theory (DFT; Erlhagen & Schoner 2002), and Leabra (Leabra; O'Reilly & Munkata 2000). These models attempt to explain mental functioning with a single system deploying a single set of operating principles. This endeavor would seem to be at odds with current dual system models in social psychology. This is because, if we examine the influential single systems models that have been developed in the broader cognitive sciences over the past 20 years, we see that these models would certainly get classified as "associative networks" rather than "rule-based systems." These models are all parallel subsymbolic networks, rather than serial symbolic logical rules. They all accommodate probabilistic soft constraints, rather than logical hard constraints. And associative systems are assumed by dual systems theory to be primitive and unsophisticated.

However, in fact, single interactive system models should be thought of as an "associative system" on steroids. The cognitive capacities of these single interactive system models far exceed those of traditional associative systems. In most dual systems models, the associative systems resemble various spreading activation networks developed in the 1970's (e.g., Collins & Loftus, 1974; cf. for a recent critical look at these models see Amodio & Ratner, 2011). However, the "connectionist revolution" in the 1980s and the "dynamic revolution" in the 1990's have advanced knowledge of what a "merely" associative system -- that is of what a parallel distributed processing network -- can do. Thus, here we coin the term "single interactive system" models of cognition to refer to these kinds of models (such as DFT, NR, and SRN, and LEABRA), with the new term reflecting the fact that these parallel distributed processing networks are no longer the associative networks of the 1970's, but self-organizing dynamical networks with sophisticated capabilities.

In particular, the connectionist revolution in the 1980's took the Collins and Loftus (1974) notion of associations between discrete symbolic concepts ("dog," "cat", etc.), and extended it to subsymbolic distributed representations (patterns of activation which may partially resemble several different concepts at once). The dynamic revolution of the 1990's took the move one step further by adding cyclic recurrent processing dynamics. Now, the layers of a connectionist network could interactively communicate with each other repeatedly over time. After these developments, it was possible to think about cognitive decisions, motor behaviors, or internal representations not as static, but as dynamically evolving over *time*.

With the new features brought on by the connectionist and dynamic revolutions, single interactive system models began to exhibit quite sophisticated behavior. Researchers have shown that feedforward connectionist networks (FN) can perform deductive syllogisms (Rogers and McClelland 2004), that simple recurrent networks (SRN) can produce rule-based language (Elman 1990; Christiansen & Chater 1999; Bechtel and Abrahamsen 1991); that normalized recurrent networks (NR), can do serial-like visual search (Spivey 2007); and that dynamic field theory (DFT) can explain the development of "symbolic thought" (Thelen & Smith 1994; Schutte & Spencer 2002). Thus, these single interactive system models possess cognitive capacities that far exceed those of the original associative networks of the 1970's, even though these models are still parallel distributed processing networks, and therefore use operating principles that conform precisely to what dual systems models in social psychology would call "System I" rather than "System II."

How much can these single interactive system models explain? A single interactive system framework has been used to explain an apparently System II phenomenon -- explicit attitudes (see Wojnowicz, Ferguson, Dale, & Spivey, 2009). In particular, a single interactive system framework has successfully described how an explicit attitude is constructed in real-time processing.[1] From this perspective (in particular, following the NR model), the active

[1] "Real-time processing" refers to the moment-to-moment timescale of information processing. The "real-time construction" of an explicit attitude would thus refer to what is happening during

construction of an explicit evaluation occurs in the following way. In the preliminary moments of processing a stimulus during an evaluation task (e.g., "do you like or dislike Black people?"), a set of informational sources simultaneously provides graded probabilistic support for multiple potential explicit decisions (see Figure 1 - still a)[2]. However, these informational sources continuously cascade probabilistic information to an integrative decision-making region (see Figure 1 - still b)[3]. The integrative decision-making region accumulates evidence for candidate decisions, forces the potential decisions to compete by way of mutual inhibition (Chelazzi & Miller 1993), and then sends top-down recurrent feedback to the informational sources, thereby updating each source's level of probabilistic support (see Figure 1 - still c). This cyclic process reiterates many times, and over recurrent cycles of activation propagation, the system gradually resolves multiple simultaneously conflicting biases, thereby settling into a finalized conclusive representation (see Figure 1 - still d). Behavioral evidence from Wojnowicz et al. (2009) suggests that this model does a good job of describing the real-time construction of an explicit attitude (for corroborating evidence see Freeman, Ambady, Rule, & Johnson 2008), using parallel distributed processing rather than discrete symbolic logical rules.

Can a single interactive system model explain evaluative readiness? After all, the single interactive system models excel precisely where the dual systems model failed. The single

---

the hundreds of milliseconds it takes for someone to respond (upon being asked) that they "like" or "dislike" something.

[2] In the context of an explicit decision to report liking versus disliking Black people, the relevant sources might include: personal memories, semantic features, subliminal evaluative conditioning, response context and future plans.

[3] Note that the integrative decision-making region is posited for convenience of the model. Neural evidence suggests integration happens at each successive step in the sensory-motor pathway; for example, motor and reward information introduce biases as far backward as primary sensory representations (Metzher et al 2006; Kay & Laurent 1999), and decision-making transformations occur as far downstream as the motor system (Zhang et al 1997; Cisek & Kalaska 2005).

interactive system models not only describe interactions (between units, layers, and/or brain regions), but interactions are actually *part and parcel* of their processing mechanisms. Moreover, the single interactive system models are neurobiologically plausible (O'Reilly 1998). Thus, single interactive system models would seem to have great potential for explaining evaluative readiness.

However, single interactive system models seem able to describe only the "slow learning system" of the brain, i.e. the posterior cortex (see O'Reilly, Braver, & Cohen 1999). That is, although single interactive system models, like the posterior cortex, are capable of processing sensory stimuli and language, it turns out that they seem poorly suited for modeling cognitive control via flexible strategic goals. To substantiate this claim, let us consider two cognitive tasks that require flexible switching between strategic goals: the Stroop task and the Wisconsin Card Sorting Task (WCST). Both of these cognitive tasks require participants to flexibly toggle their cognitive processing in accordance with actively maintained strategic goals that may switch from trial to trial. The Stoop task requires a controlled override, requiring people to categorize color word stimuli according to ink color rather than word name (as is typically done; see e.g., Engle 2002). The WCST requires participants to switch from trial to trial among categorizations according to the color, shape, texture, etc. of multi-feature stimuli (see e.g., Miyake et al 2000). In the brain, it is widely believed that good performance on these two cognitive control tasks requires the use of the prefrontal cortex to flexibly toggle between goals (Cohen, Braver & O'Reilly 1996). Computational work has arrived at the same conclusion. In particular, single interactive system networks without a specialized flexible control mechanism seem to lack the computational capacities to excel at (1) switching between goals (toggling outputs based on flexible strategic goals) and (2) transferring knowledge between goals (e.g., Rougier et al., 2005). With respect to evaluative readiness, a single interactive system model would therefore likely show deficiencies in immediately reversing its preferences for books over parties (i.e. there would be preservation), and it might flounder at transferring its knowledge to the new goal domain (it might not realize that it now dislikes not only books, but also therefore the campus library).

In sum, interactive system models have precisely the inverse capabilities of the dual systems framework: whereas the single interactive system naturally handles interactivity between component parts (e.g., it describes how an "explicit attitude" can dynamically emerge from continuously interacting component parts; see Wojnowicz et al., 2009), these models seem to be impoverished at switching between flexible strategic goals, a necessary feature of a model that explains evaluative readiness.

## III. Multiple Interacting Systems: Strategic Goals and Interactive Processing

There are brain regions whose gross fundamental processing properties differ from the posterior cortex (a slow-learning, integrative brain region which the single interactive system models described above generally most closely resemble; O'Reilly Braver & Cohen 1999). In particular, the basal ganglia have processing features enabling it to compute strategic motivational value, and the prefrontal cortex has processing features enabling it to implement flexible strategic goals (Botvinik Niv and Barto 2009; Montague et al 2004; Botvinick 2008; Koechlin & Summerfield 2007). Thus, very recent work in the computational modeling of neurobiological systems has gone beyond the single interacting system models, incorporating multiple regions with distinctive computational properties (e.g. Rougier et al. 2005; O'Reilly & Frank 2006; Botvinick, Niv & Barto 2009). These multiple interacting system models consider the single interacting system models as models of the posterior cortex, so they add additional specialized processing components reflecting the involvement of regions such as the basal ganglia and the prefrontal cortex. However, it is critical to note that these multiple interacting "systems" are not differentiated from each other in the same manner as the dual "systems" of social psychology. In contrast to the dual systems of social psychology, all of these multiple brain systems are parallel distributed processing networks, whose cognitive processing is fundamentally characterized by interaction both inside brain regions and between brain regions. What justifies the use of the term *multiple* systems is not distinct computational formats (i.e. symbolic rules versus associations) or a wall of separation between the systems (whereby communication is unclear), but rather the fact that the parallel distributed processing within these regions have distinctive specializations (in terms of neuromodulation, connectivity patterns, firing rate stability, etc.) which are functionally meaningful.[4] In this section, we will briefly identify these differences in an attempt to answer the following question: What are the specialized properties of the prefrontal cortex and basal ganglia that enable them to subserve the motivated pursuit of strategic goals? We then go on to argue for the inclusion of additional specialized processing components to the single interactive system models in order to describe the consequences for evaluative readiness (as well as for potentially numerous other psychological phenomena typically explained by dual systems).

First, the prefrontal cortex has a specialized capacity for the *active maintenance* of representations (Miller, 2000). That is, prefrontal cortical neurons can maintain goal-related activity patterns over the course of delays in working memory tasks. This capacity is not possessed by the posterior cortex. For instance, even though some neurons in temporal and posterior parietal cortex can maintain representations throughout delays, only the prefrontal cortical neurons seem to maintain their activity patterns in the face of intervening sensory stimuli. In delayed match-to-sample [4]tasks, in which participants encounter a target stimulus (e.g. a picture of a butterfly) and must signal when that target stimulus reappears after a long succession of intervening distractor stimuli (in this case, different pictures than the originally presented butterfly picture), the prefrontal cortex -- but not the posterior cortex -- maintains its activity patterns throughout the trial until the goal stimulus is found (see Miller et al 1993, 1996).

Second, the prefrontal cortex sits at the apex of a cortical hierarchy (Fuster 1997; Fuster 2004). A functional analysis of connectivity between brain regions (e.g. Stephan 1993) reveals that the prefrontal cortex has immediate privileged reciprocal access to many posterior cortical regions. The privileged hierarchical location of the prefrontal cortex enables it to provide *recurrent feedback* to the rest of the brain relevant to transient strategic goals. In particular, it is believed that the prefrontal cortex imposes top-down biases on the competition between stimuli for attention (such as a "top-down" visual search for a temporary target object amidst many other distractor objects) and between potential behaviors for motor execution (such as looking to the right before crossing the street in Britain; Desimone & Duncan 1995; Desimone 1998; Miller and Cohen 2001).

Third, the basal ganglia appear to be an "adaptive critic" (Houk, Adams, & Barto 1995), evaluating the desirability of different possible states of the world. The basal ganglia is specialized for this functionality because, unlike the posterior cortex, it contains neurons whose neurotransmitters are dopamine. Dopamine from the basal ganglia reports a reward prediction error (Bayer et al. 2005), in which case the synaptic strength between co-firing neurons is increased more than usual. Computational work suggests that the basal ganglia use this

---

[4] So, how many systems are there? Note that from a mathematical perspective, the entire brain is one dynamical system of many variables that interact with each other over time, even as those variables have different identities. From a neurobiological perspective, the number of brain systems that a research might care about would depend on her or his research interests and goals.

dopamine-enhanced learning to teach itself the motivational value of being in different cortical states -- and to learn to select actions that will produce consequences of the highest value (Montague, Hyman, & Cohen 2004, Joel, Niv & Ruppin 2002). Thus, the basal ganglia serve the role of guiding action to satisfy a person's motivational needs.

Fourth, the basal ganglia provide *dynamic gating* into the prefrontal cortex (O'Reilly & Frank 2006). *This is because* dopamine has the triple effect of (a) strengthening the currently held prefrontal representation, (b) weakening the influence of afferent sensory information, and (c) suppressing spontaneous activity (Durstewitz, Kelc, & Gunturkun 1999; Durstewitz, Seamans, & Sejnowski 2000). Thus, the basal ganglia's release of dopamine appears to stabilize prefrontal representations in the face of interfering sensory stimuli. Similarly, dopaminergic dips destabilize prefrontal cortical representations. Thus, the basal ganglia assessment of motivational value appears to determine when the prefrontal cortex switches between actively maintained goals.

Based on these four features, predominant models of executive control in computational neuroscience (Hazy, Frank, and O'Reilly 2006; Montague et al 2004) have been positing that the basal ganglia help the prefrontal cortex to "know what goals to have." Correspondingly, it seems that multiple interacting systems models are necessary to explain how people deploy strategic flexible goals. For example, Rougier et al. (2005) found that a multiple interacting systems model (which included a basal ganglia component for learning motivational value and a prefrontal cortex component for maintaining the current goal representation) outperformed a single interactive system model on the WCST and Stroop task. Moreover, even though the additional components were designed simply to implement the specialized properties of the prefrontal cortex (active maintenance extensive recurrent connectivity) and basal ganglia (adaptive criticism and dynamic gating), these components interacted to construct a fascinating emergent property: rather than representing specific features of stimuli (e.g., blue), the prefrontal cortical neurons ended up representing content-less (i.e., more abstract) dimensions. For instance, after training on the Stroop and WCST tasks, the prefrontal neurons end up representing "shape", or "color", or "size", without specifying which shape or color or size. In this way, the prefrontal cortex ended up performing the syntactic binding of the rule-based symbolic computations -- the assignment of variables to roles believed (Fodor & Pylyshyn 1988; Pinker 1997) to be the sin qua non of human language and symbolic logic thought. However, the prefrontal cortex performed these functions within a dynamic self-organizing system whose structure is parallel, distributed, and network-like. Thus, the multiple interacting systems models can implement "rule-based behavior" without losing interactivity or its subsymbolic, parallel, and distributed processing properties.

We turn now, finally, to addressing our original problem: *how* can we explain evaluative readiness? Remember that our theoretical question was: how do flexible strategic goals influence implicit attitudes? To address this question, let us briefly revisit how implicit attitudes are measured in the evaluative readiness literature. A person's evaluation of "cigarettes" for example is typically measured based on reaction times to positive words such as "good" after being primed with a stimulus word like "cigarette." From the perspective of most connectionist networks with distributed representations, reaction times are a proxy for proximity in state space (how long it takes the mind to transition between distributed patterns) (e.g., Cree, McRae, & McNorgan 1999). In other words, more similar distributed patterns would show stronger priming. Thus, when a person wants a cigarette, the distributed pattern for "cigarette" and the distributed pattern for positive concepts like "good"[5] should look more similar than when a person doesn't want a cigarette. This means that a person's currently operating strategic goal should modulate the representation of concepts![6] How might this be happening?

We may shed light on this question if we look more deeply at how the prefrontal cortex was affecting the posterior cortex in the Rougier et al. (2005) simulation. Their posterior cortex represented each stimulus inside a 145-unit layer. Thus, a medium-sized blue square stimulus with a given texture and location would have a 145-dimensional representation. (That is, the stimulus would be represented by a particular pattern of neural firing rates over 145 neuronal units, and thus would be captured by a point in a 145-dimensional Euclidean space). However, when the prefrontal cortex in that model was maintaining the strategic goal relevant to the current trial of the WCST (e.g. to categorize by color), it seemed to be effectively projecting the posterior cortical representation onto a smaller dimensional subspace (e.g. 19-dimensional representation).[7] The smaller-dimensional representation would preserve the posterior cortical information relevant for the current goal -- e.g. the stimulus' color -- while discarding information irrelevant to the current goal -- e.g. the stimulus' size, shape, texture, and location. But as the agent's strategic goals flexibly changed over the course of the task, the prefrontal cortex flexibly shifted its projections. That is, the prefrontal cortex component, through dynamic gating from the basal ganglia, flexibly switched between different lower-dimensional

[5] Or positively valenced concepts, like sunshine and puppies.
[6] For a similar point with emotions, see Niedenthal, Halberstadt, & Innes-Ker 1999.
[7] Temporarily projecting a 145-dimensional space onto a 19-dimensional subspace would mean that, for the time being, only 19 of those dimensions are "used." For a simpler example, consider a bivariate data set. To "project" points in the x,y plane onto the line y=2x+3 would mean that every point in x,y space would be reassigned to the closest corresponding point on the line. Moments later, the same point might be projected onto a different line.

representations (color or shape or location) of the same high-dimensional stimuli.[8] In fact, these low-dimensional projections seem to be precisely what are responsible for the effective performance of the multiple systems model in Rougier et al (2005), who found strong correlations between (a) the model's performance on strategic flexible goal tasks and (b) how well the prefrontal cortex component learned orthogonal dimensions that it could feedback into the posterior cortex.

Thus, we argue here that the current best theoretical framework for understanding evaluative readiness seems to be through a "dynamic projections" perspective. That is, the prefrontal cortex seems to be adaptively selecting a lower-dimensional subspace upon which to project the high-dimensional representations of the posterior cortex. The selection of the subspace depends upon the currently operating strategic goal. When the prefrontal cortex chooses a different strategic goal, it actively maintains a different firing pattern, which thereby projects the posterior cortical representations onto a different subspace. This projection mechanism would explain evaluative readiness, because priming times (reflecting evaluations) would depend upon the projection onto various lower-dimensional subspaces.

For a simple cartoon example, imagine a posterior cortex with only three neurons. Then, the "high-dimensional" posterior cortical representations for semantic concepts would have 3 dimensions, and lower-dimensional projections would have fewer dimensions (perhaps 2). Figure 2 shows some hypothetical distributed representations for libraries, good, and parties in the hypothetical three-neuron posterior cortex. Figure 3 depicts these same three hypothetical representations in three-dimensional Euclidean space, where (x,y,z) = (-4,0,10) for "parties", (4,

---

[8] A familiar example of a projection onto a low-dimensional subspace would be the use of factor analysis to find the low-dimensional factors (embedded within a high-dimensional space) which can capture the maximum variance of a data set.

1,9) for "libraries", and (4,1,1) for "good." Whether the mental representations for "libraries" and "good things" are close to each other or far depends on how the prefrontal cortex projects the pattern-formation dynamics onto subspaces. If the projection occurs along the plane Q, then "libraries" primes "good" very strongly, much more strongly than "parties." If the projection occurs along the plane P, "libraries" would not prime "good" very strongly, especially in comparison to "parties."

The main point is that recent multiple interacting systems models suggest that -- based on the distinguishing properties of the prefrontal cortex and basal ganglia -- dimensionality projections are what allows the prefrontal cortex to achieve flexible cognitive control in accordance with current motivations. In the case of the Rougier et al. (2005) model of the WCST, the prefrontal cortex provided a mechanism through which the model could flexibly shift around the similarity groupings of blue squares vs. green squares vs. green circles, in accordance with task demands. We construe evaluative readiness in exactly the same way -- the prefrontal cortex provides a mechanism through which the posterior cortex flexibly shifts around evaluations of books and libraries vs. parties and beers depending on current goals.

## Conclusion to Section 2

In this section, we sketched out a theory for *how* evaluative readiness works. We suggested that two popular classes of information processing models-- dual systems models (from social psychology) and single interacting system models (from cognitive science) -- would be inadequate by themselves to explain evaluative readiness. In particular, dual systems models seem to require the strict separation of strategic goals and implicit attitudes, failing to describe a mechanism by which these two psychological constructs could interactively influence each other. On the other hand, single interacting system models are rooted in rich interactivity, but flexible strategic goals are nowhere to be found in these models. Thus, we have grounded our considerations within a more contemporary, hybridized class of models for flexible cognitive control, which we call (in contrast to the other two classes of models) multiple interacting systems models. These models describe rich interacting processing within and between mental systems, and seem well-suited for implementing flexible strategic goals. These models suggest that the phenomenon of evaluative readiness happens through motivated projections. In particular, current strategic goals (which chosen by the basal ganglia and actively maintained in the prefrontal cortex) cause a projection of high-dimensional mental representations in the posterior cortex onto lower-dimensional subspaces. These projections onto lower-dimensional subspaces can alter which concepts are "near" other concepts in state space (i.e. it changes how similar their distributed patterns are), thereby producing goal-moderated categorizations (as in the WCST) and goal-moderated evaluations (as in evaluative priming). These multiple interacting systems models might be useful in explaining other psychological phenomena that have typically been explained by traditional dual systems models. This approach to understanding evaluative readiness joins other recent attempts to explain and model social psychological phenomena using social neuroscience theory and findings (e.g., see Amodio, 2010; Todorov, Fiske, & Prentice, 2011).

**References**

Aarts, H., Diksterhuis, A., & De Vries, P. (2001). On the psychology of drinking: Being thirsty and perceptually ready. British Journal of Psychology, 92, 631-642.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

Alberts, J.R. (2007). Huddling by rat pups: Ontogeny of individual and group behavior. Developmental Psychobiology, 49, 22-32.

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition, 28*, 695-716.

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

Atkinson, J. (1964). An introduction to motivation. Oxford, England: Van Nostrand.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction, and reinforcement learning in the basal ganglia. Progress in Neurobiology, 71, 439-473.

Bargh, J. A. (1990). Auto-motives: Preconscious determinants of social interaction. In E. T. Higgins, & R. M. Sorrentino (Eds.), Handbook of motivation and cognition: Foundations of social behavior (Vol. 2., pp. 93-130). New York, NY, USA: The Guilford Press.

Barto, A.G. & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement leartning. Discrete Event Dyanmical Systems, 13: 341-379.

Bechtel, W. & Abrahamsen, A. (1991). Connectionism and the mind: An introduction to parallel processing in networks. Basil Blackwell.

Bever, T.G. (1970). The cognitive basis for linguistic structures. Cognition and the Development of Language, 279: 352.

Botvinick, M.M. (2008). Hierarchical models of behavior and prefrontal function. Trends in Cognitive Sciences, 12 201-208.

Botvinick, M.M., Niv, Y., & Barto, A. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. Cognition, 113, 262-280.

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

Camazine, S., Deneubourg, J.L., Franks, N.R. Sneyd, J., Theraula, G. & Bonabeau, E. (2003). Self-Organizational in Biological Systems. Princeton University Press.

Carver, C. S., & Scheier, M. F. (1981). Attention and Self-regulation: A control-theory approach to human behavior. New York: Springer-Verlag.

Carver, C. S. & Scheier, M. F. (1998). On the self-regulation of behavior. New York: Cambridge University Press.

Chelazzi, L. & Miller, E.K. (1993). A neural basis for visual search in inferior temporal cortex. Nature 345, 347.

Churchland, P.S. & Sejnowski T.J. (1994). The Computational Brain. Boston: MIT Press.

Christiansen, M.H. & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. Cognitive Science, 23, 157-205.

Cisek, P. & Kalaska. J.F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron, 45: 801-814

Cree, G.S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. Cognitive Science: A Multidisciplinary Journal, 23, 371-414.

Cunningham, W.A. & Zelazo, P.D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. Trends in Cognitive Sciences, 11, 97-104.

Custers, R., & Aarts, H. (2005). Positive Affect as Implicit Motivator: On the Nonconscious Operation of Behavioral goals. Journal of Personality and Social Psychology, 89(2), 129-142.

Dale, R., Kehoe, C., & Spivey M.J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. Memory and Cognition 35, 15-28.

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

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. Psychological Bulletin*, 127*, 853–869.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum.

Desimone, R. (1998). Visual attention mediated by biased competition in the extrastriate visual cortex. Philosophical Transactions of the Royal Society B: Biological Sciences, 353: 1245.

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18, 193-222.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. Journal of Personality and Social Psychology, 86, 345-355.

Dovidio, Kawakami, & Gaertner (2002). Implicit and explicit prejudice and interracial interaction. Journal of Personality and Social Psychology, 82, 52-58.

Durstewitz, D., Kelc, M. & Gunturkun, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. Journal of Neuroscience, 19, 2807-2822.

Dustetwitz, D., Seamans, J.K., & Sejnowski, T.J. (2000). Neurocomputational models of working memory. Nature Neuroscience, 3, 1184-1191.

Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Engle, R.W. (2002) Working memory capacity as executive attention. Current Directions in Psychological Science, 11, 19.

Erlhagen & Schoner (2002). Dynamic field theory of movement preparation. Psychological Review, 109, 545.

Farmer, T.A., Anderson, S.E., & Spivey, M.J. (2007). Gradiency and visual context in syntactic garden-paths. J. Mem. Lang. 57, 570-595.

Fazio, RH. (1990). Multiple processes by which attitudes guide behavior: the mode model as an integrative framework. In: Advances in experimental social psychology, 23, 75-109.

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. (2001). On the automatic activation of associated evaluations: An overview. Cognition and Emotion, 14, 1-27.

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.

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. Journal of Personality and Social Psychology, 50, 229-238.

Ferguson, M. J. (in press). On becoming ready to pursue a goal you don't know you have: Effects of nonconscious goals on evaluative readiness. Journal of Personality and Social Psychology.

Ferguson, M. J. (2007). On the automatic evaluation of end-states. Journal of Personality and Social Psychology, 92, 596-611.

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

Ferguson, M. J., & Bargh, J. A. (2008). Evaluative readiness: The motivational nature of automatic evaluation. Invited chapter in A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation*. Lawrence Erlbaum.

Ferguson, M. J., & Porter, S. C. (in press). Goals and attitudes: A social cognitive perspective. Invited chapter to appear in G. Moscowitz & H. Grant (Eds.), *The Handbook of Goals*. New York: Guilford Press.

Fishbach, A., & Ferguson, M. J. (2007). The goal construct in social psychology. Invited chapter in A. W. Kruglanski & E. T. Higgins & (Eds.). Social psychology: Handbook of basic principles (pp. 490-515). Volume II. New York: Guilford Press.

Fishbach, A., Friedman, R., & Kruglanski, A. W. (2003). Leading us not unto temptation: Momentary allurements elicit automatic goal activation. Journal of Personality and Social Psychology, 84, 296-309.

Fishbach, A., Shah, J. Y., & Kruglanski, A. W. (2004). Emotional Transfer in Goal Systems. Journal of Experimental Social Psychology, 40, 723-738.

Fishbach, A., Zhang, Y., & Trope, Y. (2010). Counteractive evaluation: Asymmetric shifts in the implicit value of conflicting motivations. *Journal of Experimental Social Psychology, 46*, 29-38.

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

Forster, J., Liberman, N., & Higgins, E. (2005). Accessibility from active and fulfilled goals. Journal of Experimental Social Psychology, 41(3), 220-239.

Freeman, J.B., Ambady, N., Rule, N.O., & Johnson, K.L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. Journal of Experimental Psychology: General, 137, 673-690.

Fujita, K., Trope, Y., Liberman, N., & Levi-Sagi, M. (2006). Construal levels and self-control. Journal of Personality and Social Psychology, 90, 351-367.

Fuster, J.M. (2004). Upper processing stages of the perception-action cycle. Trends in Cognitive Sciences, 8, 143-145.

Gallagher, M. & Holland, P.C. (1994). The amygdala complex: Multiple roles in associative learning and attention. Proceedings of the National Academy of Sciences, 91, 11771-11776.

Gawronski & Bodenhausen (2006). Associative and Propositional Processes in Evaluation: An Integrative Review of Implicit and Explicit Attitude Change. Psychological Bulletin, 132, 692-731.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), Handbook of motivation and cognition: Foundations of social behavior (Vol. 2, pp. 53-92). New York: Guilford Press.

Greenwald, A.G. & Nosek, B.A. (2008). Attitudinal dissociation: What does it mean? In: Attitudes: Insights from the new implicit measure. Citeseer Press.

Hazy, T.E., Frank, M.J., & O'Reilly R.C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. Philosophical Transactions of the Royal Society B: Biological Sciences, 362, 1601-1613.

Higgins, E. T. (1997). Beyond pleasure and pain. American Psychologist, 52, 1280-1300.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 30, pp. 1-46). New York: Academic Press.

Higgins, E. T., Shah, J., & Friedman, R. (1997). Emotional responses to goal attainment: Strength of regulatory focus as moderator. Journal of Personality & Social Psychology, 72, 515-525.

Houk, J.C., Adams, J.L., & Barto, A.G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Models of information processing in the basal ganglia.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: New anatomical and computational perspectives. Neural Networks, 15, 535-547.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Giloivch, D. Griffin and D. Kahneman (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment (pp. 49-81).

Kay, L.M. & Laurent, G. (1999) Odor- and context-dependent modulation of mitral cell activity in behaving rats. Nature Neuroscience, 2, 1003-1009.

Koechlin, E. & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. Trends in Cognitive Sciences, 11, 229-235.

Koob, G.F. & Bloom, F.E. (1988). Cellular and molecular mechanisms of drug dependence. Science, 242, 715-723.

Kruglanski, A. W. (1996). Goals as knowledge structures. In P. M. Gollwitzer and J. A. Bargh (Eds.), The psychology of action: Linking cognition and motivation to behavior (pp. 599-618). New York: Guilford.

Lisker, L. (1986). Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. Language and speech, 29, 3-11.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice Hall.

Mates, B. (1989). Philosophy of Leibniz: Metaphysics and Language. Oxford University Press.

McConnell, J. & Leibold (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. Journal of Experimental Social Psychology, 37, 435-442.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 276, 746-748.

Metzher, R.R., Greene, N.T., Porter K.K. & Groh, J.M. (2006). Effects of reward and behavioral context on neural activity in the primate inferior colliculus, The Journal of Neuroscience, 28, 7468.

Miller, E.K., Ericson, C.A. &Desimone, R. (1996). Neural mechanisms of visual working memory in the prefrontal cortex of the macaque. Journal of Neuroscience, 16, 5154-5167.

Miller, E.K. & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annual review of neuroscience, 24, 167-202.

Mischel, W., Cantor, N., & Feldman, S. (1996). Principles of self-regulation: The nature of willpower and self-control. In E. T. Higgins & A. Kruglanski (Eds.), Social psychology: Handbook of basic principles (pp. 329-360). New York: Guilford Press.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. Science, 244, 933-938.

Miyake, A. Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager T.D. (2000). The unity and diversity of executive functions and their contributions to complex. Cognitive Psychology, 41, 49-100.

Montague, P.R., Dayan, P. & Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. Journal of Neuroscience, 16, 1936-1947.

Montague, P.R., Hyman, S.E., & Cohen, J.D. (2004). Computational roles for dopamine in behavioral control. Nature, 431, 760-767.

Moskowitz, G. B. (2002). Preconscious effects of temporary goals on attention. Journal of Experimental Social Psychology, 38, 397-404.

O'Reilly, R.C. (1998). Six principles for biologically based computational models of cortical cognition. Trends in Cognitive Science 2, 455-462.

O'Reilly, R.C., Braver, T.S., & Cohen, J.D. (1999). A biologically-based computational model of working memory. In: Models of working memory: Mechanisms of active maintenance and executive control. Citeseer Press.

O'Reilly, R.C. & Frank, M.J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Computation, 18, 283-328.

O'Reilly, R.C. & Munkata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. The MIT Press.

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

Petty, R.E. and Brinol, P. and DeMarree, K.G. (2007). The Meta-Cognitive Model (MCM) of Attitudes: Implications for Attitude Measurement, Change, and Strength. Social Cognition, 25, 657-686.

Petty, R. E., Fazio, R. H., & Briñol, P. (in press). Attitudes: Insights from the new wave of implicit measures. Mahwah, NJ: Erlbaum.

Pinker, S. (1997). How the Mind Works. NY: Norton Press.

Pouget, A., Dayan, P. & Zemel, R. (2000). Information processing with population codes. Nature Reviews Neuroscience,1, 125-132.

Redgrave, P., Prescott, T.J., Gurney K (1999). The basal ganglia: A vertebrate solution to the selection problem? Neuroscience Oxford, 89, 1009-1024.

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

Romo, R. & Schultz, W. (1990). Dopamine neurons of the monkey midbrain: contingencies of response to active touch during self-initiated arm-movements. Journal of Neurophysiology, 63, 592-606.

Rougier, N.P., Noelle, D.C., Braver, T.S., Cohen, J.D., & O'Reilly, R.C. Prefrontal cortex and flexible cognitive control: Rules without symbols. Proceedings of the National Academy of Sciences of the United States of America, 102: 7338.

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

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

Schultz, W. & Romo, R. (1990). Dopamine neurons of the monkey midbrain: contingencies of response to stimuli eliciting immediate behavioral reactions. Journal of Neurophysiology, 63, 607-624.

Seibt, B., Häfner, M. & Deutsch, R. (2007). Prepared to eat: How immediate affective and motivational responses to food cues are influenced by food deprivation. European Journal of Social Psychology.

Schutte, A.R., Spencer, J.P. & Schoner, G. (2002). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. Child Development, 74: 1393 -1417.

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. Journal of Personality and Social Psychology, 83, 1261-1280.

Shah, J., & Gardner, W. (2007). Handbook of motivational science. New York, NY: Guilford Press.

Shah, J. Y., & Kruglanski, A. W. (2002). Priming against your will: How goal pursuit is affected by accessible alternatives. Journal of Experimental Social Psychology, 38, 368-382.

Shah, J.Y., & Kruglanski, A.W. (2003). When opportunity knocks: Bottom-up priming of goals by means and its effects on self-regulation. Journal of Personality and Social Psychology, 84, 1109-1122.

Sherman, S. J., Presson, C. C., Chassin, L., Rose, J. S., & Koch, K. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. Journal of Social and Clinical Psychology, 22, 13-39.

Smith, E.R. & Conrey, F.R. (2007). Mental representations are states, not things: Implications for implicit and explicit measurements. Implicit measures of attitudes, 247-264.

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.

Stephan, K.E., Hilgetag, C.C., Burns, G.A.P.C, O'Neill, M.A., Young, M.P. & Kotter, R. (1993). Computational analysis of functional connectivity between areas of primate cerebral cortex. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 355, 111.

Sutton, R.S. & Barto, A.G. (1998). Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.

Spivey, M.J. (2007). The Continuity of Mind. Oxford: Oxford University Press.

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

Thelen, E. & Smith, L.B. (1994). A dynamic systems approach to the development of cognition and action. MIT Press.

Todorov, A., Fiske, S. T., & Prentice, D. (Eds.). (2011). *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. Oxford University Press.

Tononi, G., Sporns, O., & Edelman, G.M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. Proceedings of the National Academy of Sciences of the United States, 91, 5033-5044.

Uleman, J.S. (1999). Spontaneous versus intentional inferences in impression formation. In: Dual-process Theories in Social Psychology, 141-160.

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

Wittenbrink, B., & Schwarz, N. (2007). Implicit measures of attitudes. New York, NY: Guilford Press.

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

Yeh, W., & Barsalou, L.W. (2006). The situated nature of concepts. American Journal of Psychology, 119, 349-384

Zajonc, R.B. (1980). Feeling and thinking: preferences need no inferences. American psychologist, 35, 151.

Zhang, J., Riehle A., Requin J., & Kornblum S. (1997). Dynamics of single neuron activity in monkey primary motor cortex related to sensorimotor transformation, 17, 2227.