Title: "Hot" cognition and dual systems: Introduction, criticisms, and ways forward

Authors: Thomas E. Gladwin *, a, b
          Bernd Figner *, a, c

a Radboud University, Nijmegen, The Netherlands.
b University of Amsterdam, Amsterdam, The Netherlands
c Center for the Decision Sciences, Columbia University

* Corresponding authors. Email: thomas.gladwin@gmail.com, b.figner@psych.ru.nl

Acknowledgements: We would like to thank Karin Roelofs and Vanessa van Ast for their helpful feedback.

This is an Author’s Accepted Manuscript of an article published in 2014 in Frontiers of Cognitive Psychology Series: Neuroeconomics, Judgment and Decision Making, Eds Wilhelms E, Reyna VF, available at http://www.amazon.com/Neuroeconomics-Judgment-Frontiers-Cognitive-Psychology/dp/1848726600/ref=tmm_pap_title_0.
Abstract

Models distinguishing two types of processes or systems—typically one more automatic and/or affective-motivational, one more controlled and/or calculating-deliberative—are widespread in psychological science. However, such dual-process (or dual-system) models suffer from various problems and have been substantially criticized recently. In this chapter, we discuss these types of models, attempt to clarify terminology, discuss recent critiques at both empirical and theoretical levels, and suggest a more mechanistic explanation grounded in physiology and reinforcement learning of what makes "hot" processes hot. We discuss success stories and challenges related to these types of models in two illustrative fields, addiction and adolescent risk taking. Finally, we outline the basic ideas behind our R3 model—a reprocessing model grounded in reinforcement learning that conceptualizes levels of reflectivity as emergent states of one single system, rather than a separate process or system—as a possible way forward to address and overcome problems of dual-process models.
Models distinguishing two types of different processes or "systems" are prominent and widespread in many fields of psychological science. However, they recently have been substantially criticized and challenged. In this chapter, we focus on so-called dual-process or dual-system models that differentiate between more automatic (often "hot" emotional-affective) versus more controlled (often "cold" cognitive-deliberative) processes. We start out with an attempt to describe and clarify different terminologies, including a clarification of the temperature metaphor of "hotness versus coldness." We then propose to ground and decompose the notion of "hotness" in emotion-relevant basic biological processes of the autonomic nervous system and incentive salience. Extending the scope, we then focus on two types of dual-process or dual-system models, discussing both their strengths as well as shortcomings. Finally, we suggest a diagnosis of the current state of affairs and propose possibly more fruitful directions for future research and theory-forming. As part of this, we briefly describe our R3 model, a novel model of reflectivity that here serves as a proof-of-principle thought-experiment to address several shortcomings of existing dual-process and dual-system models.

The temperature metaphor
The temperature metaphors of "hot" versus "cold" or "cool" phenomena are used widely in different forms in psychological science, typically referring to a differential involvement of processes related to affect and/or motivation on the one hand versus more controlled and/or cognitive processes on the other hand. For example, the terms "hot" versus "cold" cognition refer to cognitive processes that are relatively affect-charged versus affect-free, respectively (Abelson, 1963). However, the term "hot cognition" has been used to refer to both affective processes themselves (such as emotions or feelings) as well as phenomena such as emotional appraisal, which arguably could be considered a cognitive process (albeit one crucially relevant for affect
and emotion). Thus, the mapping between hot vs. cold cognition on the one hand and affect vs. cognition on the other hand is somewhat ambiguous. Similarly to Abelson, an influential paper by Metcalfe and Mischel (1999) used the temperature metaphor to differentiate between a hot versus a cool system. The former refers to a system encompassing affective and motivational processes and the latter refers to a system encompassing more cognitive-deliberative processes. Grounded in a developmental perspective, Metcalfe and Mischel characterize the hot system as being under stimulus control, emotional, fast, reflexive, relatively simple, and developing relatively early during human ontogeny. In contrast, the cool system is characterized by self-control, encompasses cognitive processes, and is comparatively slow, reflective, complex, and develops relatively late during ontogeny. These and similar characteristics have been widely used to describe the two types of processes or systems often distinguished in the literature (discussed later in this chapter).

Another, somewhat different use of referring to hot vs. cold processes—which appears to be particularly prominent in research on ADHD and in developmental psychology—is the distinction between hot versus cool (or cold) executive functions (EF) (Prencipe et al., 2011; Van den Wildenberg & Crone, 2005): The term "cold EF" is used to describe executive functions conceptualized as lacking an affective component, such as working memory and inhibition; accordingly, tasks such as backwards digit span or Color Word Stroop are typically used to assess cold EF. The main neural substrate serving cold EF is assumed to be the dorsolateral prefrontal cortex (DLPFC). In contrast, the term hot EF is typically used to describe EF that involve an affective or motivational component. Typical tasks used in the literature to assess hot EF are often decision-making tasks, such as intertemporal choice tasks (also referred to as delay discounting tasks) or risky decision making tasks, in particular the Iowa Gambling Task.
(Bechara, Damasio, Damasio, & Anderson, 1994) and variants thereof like the Hungry Donkey task (Geurts, van der Oord, & Crone, 2006; Hongwanishkul, Happaney, Lee, & Zelazo, 2005). The main neural substrates serving hot EF are assumed to be more ventral and medial regions of the prefrontal cortex, including the orbitofrontal cortex.

The idea of using performance in decision-making tasks to operationalize hot EF, however, is likely contributing to the existing variations and inconsistencies in terminology: Decision making itself is assumed to typically involve various processes, including cognitive and control processes and affective processes. For example the "risks-as-feeling" hypothesis (Loewenstein et al., 2001) posits that, when faced with a risky choice, we not only evaluate the available options via more cognitive and deliberative processes, but (i) also have emotional responses to characteristics of the choice options (such as their "risk," i.e., their outcome variability) and (ii) that these emotional responses may have a stronger impact on choice than the more cognitive evaluations. Another role of affective processes in decision making has been termed "common currency" (Cabanac, 1992; Figner & Weber, 2011; Levy & Glimcher, 2012; Peters, Västfjäll, Gärling, & Slovic, 2006): The idea is that emotions serve as the basis to evaluate and choose among attributes, goods, and outcomes that otherwise would be incommensurable. Further complicating matters, processes of self-control to resist "hot" temptations have been shown, via experimental interference with neural processing using non-invasive brain stimulation techniques, to causally involve the DLPFC (Figner et al., 2010; Knoch et al., 2006), which, in the framework of hot vs. cold EF would be the main neural substrate for cold, not hot, EF. Finally, decision making very reliably involves neural processes in the ventromedial and orbital PFC (Carter, Meyer, & Huettel, 2010), seemingly consistent with the idea of hot EF and its neural substrates. However, in the decision neuroscience and neuroeconomics literatures, the involvement of these regions is
typically not described as hot EF, but as reflecting the subjective value of the presented choice options, reflecting an evaluation process that likely is at least partly affective in nature, consistent with the "common currency" idea.

At the very least these points show that using decision-making tasks as measures for hot EF can be problematic, as such tasks cannot not be considered "process-pure" measures of the underlying characteristics or processes specified by a given researcher’s definition of hot EF. This touches on a more worrisome point: the differential terminology and frameworks have the potential to cause substantial misunderstanding and confusion among scientists. First, about what is being measured and observed; second, about how the results should be interpreted with respect to the involved psychological and neural processes; and third, about what we can learn from the studies when we try to integrate results and insights across different fields, thus making it even more challenging to build more overarching models that cover a wider range of empirical work.

In the spirit of full disclosure, we should declare that one of the authors of this chapter also contributed to the already complex and confusing use of temperature metaphors, as his co-authors and he named two versions of his risky decision-making task "hot" and "cold," respectively; namely the hot and cold Columbia Card Task (Figner, Mackinlay, Wilkening, & Weber, 2009a): The hot version was designed to involve substantially affective-motivational decision processes, while the cold version was designed to involve predominantly cognitive-deliberative processes. Again, both tasks naturally involve affective, deliberative, and self-control processes, however, to a differential extent, and this differential involvement has inspired the use of hot and cold in the task name.
To sum up these introductory remarks, while it seems that the use of temperature metaphors is quite popular in psychological science to refer to two different types of processes or systems, these terms are not always used with the same meaning in mind, and the distinctions and implied brain regions do not always neatly line up. This is already sufficient reason to be careful to avoid creating more confusion than enlightenment when using connotations associated with hot and cold to explain psychological phenomena. But a more important goal is that we should start to lay a more solid foundation for these concepts; we suggest a possible way to do this in the next section.

**From “heat” to autonomic responses and incentive salience**

The “heat” in temperature metaphors appeals to the subjective experience of arousal and emotion—the memory, experience, and anticipation of pounding hearts, sweaty palms, rapid breathing, and so on. Such basic biological responses are patterns that prepare the body to deal with events related to survival and procreation, that is, to support fitness-enhancing behavior essential in evolution: defensive reactions to threat, or appetitive responses to attractive stimuli. For example, when confronted with threatening stimuli a "freeze"-response may occur with characteristic physiological changes such as decreased heart rate and heart rate variability, increased skin conductance, and reduced body sway (Bracha, 2004; Dalton, Kalin, Grist, & Davidson, 2005; Jarvik & Russell, 1979; Roelofs, Hagenaars, & Stins, 2010). To some authors, it is introspectively obvious that the representation of such changes in the autonomic nervous system (ANS) is the defining feature of emotion. For instance, William James (1884), p. 451, wrote: *"What kind of an emotion of fear would be left, if the feelings neither of quickened heart-beats nor of shallow breathing, neither of trembling lips nor of weakened limbs, neither of goose-flesh nor of visceral stirrings, were present, it is quite impossible to think."* Accordingly, in the James-Lange theory of emotion, and its subsequent scientific lineage (Reisenzein, Meyer, &
Schutzwohl, 1995), an emotion is the feeling of stimulus-evoked bodily, especially visceral, states. No consensus has been reached on the precise relationship between emotion and ANS activity, although studies, using increasingly sophisticated methods, show a close, and perhaps even emotion-specific, coupling between emotions and patterns of ANS responses over various physiological measures (Collet, Vernet-Maury, Delhomme, & Dittmar, 1997; Kreibig, 2010; Stephens, Christie, & Friedman, 2010). Here, we note that from a modern neuroscientific perspective, the “autonomic” nervous system, far from being independent from the central nervous system, can be traced up to the cortex (Kreibig, 2010). A functional unit of brain regions that has been termed the Central Autonomic Network—which includes regions with notably strong links to emotion such as the periaqueductal gray, insula, central amygdala, anterior cingulate, and ventromedial prefrontal cortex—send output to efferent regions of the medulla that affect sympathetic and parasympathetic ANS activity (Benarroch, 1993; Cersosimo & Benarroch, 2013; Napadow, Dhond, Conti, & Makris, 2008; Thayer & Lane, 2000).

Note that a necessary role of the central representation of ANS responses in emotion does not necessarily require a sequence of events actually involving the periphery. As with other central neuronal structures and processes, their function and meaning is ultimately derived from the biologically relevant environment. For instance, occipital regions represent visual stimuli, and motor regions represent movements; therefore, processes that involve these regions can be interpreted in terms of vision and movement, respectively, even in the absence of visual stimulation or the actual execution of muscle contraction. In our case, we are considering neurons that represent physiological states related to emotional responses. For instance, James’ introspective feelings of emotions presumably most directly involve these central neural representations, rather than physiological responses that are usually connected with them. Thus,
claiming that patterns of ANS activation play a core role in defining emotion refers to the consistent involvement of central neurons of which the function or representation is defined by ANS patterns. The function these neurons have, or at least would initially attempt to fulfill, in navigating the world remains the same even when they, for instance in surgical animal research or in disease, are physically disconnected from the ANS; just like visual neurons could be disconnected from the eyes or motor neurons disconnected from the limbs.

From this perspective, a definition of “hot” stimuli must include reference to “hot” physiological states, evoked due to those stimuli having evolutionary significance directly or to being linked to such significance via conditioning. Correspondingly, emotion regulation (a “hot” executive function in the distinction between hot and cold EF) would be understood to a large extent as having physiological regulation as its final outcome. Intuitively, this is what we subjectively perceive as successful regulation: when we “control,” usually meaning ”down-regulate,” our emotions, our breathing slows, our heart rate goes down, our blood pressure drops; and presumably other visceral sensations change, which we may be less able to consciously identify.

The above focuses on the visceral part of emotion, which, although to some authors the defining feature of emotion, is clearly not the whole story of hot cognition. A foundation for understanding the more central, i.e., neural and psychological, components of “hotness” is Incentive Salience Theory (Berridge & Robinson, 1998; LeDoux, 2012; Robinson & Berridge, 1993). This is a prominent theory on the role of dopamine in reward learning, which states that stimuli and cues that are associated with mesolimbic dopamine release acquire incentive salience: the ability to attract attention and act as a reward for behavior. While usually phrased in positive terms, incentive salience appears also applicable to behavior and responses based on a negative
"wanting," i.e., associations with the tendency to escape or avoid aversive stimuli (although it is unclear whether the neural processes of such appetitive and aversive incentive saliency would overlap). The ability to evoke behavior aimed towards a goal involving the stimulus (termed “wanting”), measured via choice behavior, can be distinguished from the hedonic effects of a stimulus (termed “liking”), measured via observable reflections of pleasure or dislike such as facial expression and taste reactivity. One line of evidence for relating dopamine to incentive salience is that dopamine release is associated with the initiation of approach behavior rather than consumption. Further, animal research shows that depletion of dopamine in the mesolimbic system does not affect either “liking” or the learning of hedonic associations (Berridge & Robinson, 1998). An important aspect of the theory is that it explains flexible goal-directed behavior in a mechanistic fashion: What is learned via the acquisition of incentive salience is not a rigid motoric stimulus–response association, or the hedonic value of a stimulus, but the incentive value of a stimulus. Incentive value can be operationalized as the amount of work a stimulus evokes, reflecting how much effort and cost approaching or avoiding the stimulus is “worth.” Incentive salience could alternatively be described as a stimulus-goal association, or as a stimulus-dependent action-outcome association (Dickinson & Balleine, 2011). Such descriptions appear to be in line with studies showing a relationship between activation in the mesolimbic system, in particular the ventral striatum, and learning how to respond in such a way as to optimize feedback (Bunge, Burrows, & Wagner, 2004; Day & Carelli, 2007; Delgado, Miller, Inati, & Phelps, 2005; O’Doherty, Hampton, & Kim, 2007; Seger, 2008; Vink, Pas, Bijleveld, Custers, & Gladwin, 2013). Ideally, the incentive value of a stimulus encodes whether the stimulus or outcome predicts, perhaps indirectly, an evolutionarily relevant event, such that it can function as reward or punishment, although the system is not perfect, as evidenced by addictive drugs, which are thought to tap relatively directly into this dopaminergic system, but without
having the associated evolutionary fitness advantage. Once the goal to acquire the incentive is activated, the actions needed to achieve that goal will be determined and recruited via other processes, depending on the context and prior learning (Robbins & Everitt, 1999; Tiffany, 1990).

Addiction provides an example of incentive salience gone awry. Addictive drugs or behaviors such as gambling have been proposed to cause incentive sensitization: Repeated use leads to an increase in the mesolimbic responses underlying incentive salience, causing drugs to become increasingly “wanted” stimuli (Robinson & Berridge, 2008). In animals, exposure to drugs results in a wide range of conditioning effects including self-administration acquisition, conditioned place preference, the amount of work an animal will perform for a drug, conditioned reinforcement, and Pavlovian conditioned approach and Pavlovian instrumental transfer (the last two providing more mechanistic models how "hot" processes can conflict with and "hijack" goal-directed "cold" processes) (for details on effects of drug exposure, see the review by Robinson and Berridge, 2008). In humans, such research is rare, but in a placebo-controlled PET study, stimuli that have been associated with amphetamine delivery acquire the ability to release dopamine in the striatum (Boileau et al., 2007), just as the sound of a bell acquired the ability to cause salivation in Pavlov’s dogs. Drug-related incentive salience can also be detected using behavioral methods. For instance, individuals who drink heavily but not in a clinically problematic way show an attentional bias towards alcohol cues (Field, Mogg, Zetteler, & Bradley, 2004; Townshend & Duka, 2007). In alcohol-dependent individuals, a more complex pattern has been shown, with initial orienting towards alcohol cues followed by attentional disengagement (Noël et al., 2006; Vollstädt-Klein, Loeber, von der Goltz, Mann, & Kiefer, 2009). Approach-avoidance biases, reflecting a stronger tendency to approach than avoid alcohol cues (although this is controversial), have also been found. One task used to assess such biases is
the Approach Avoidance Task or AAT (Enter, Colzato, & Roelofs, 2012; Rinck & Becker, 2007). In this canonical version of this task, subjects are confronted with stimuli drawn from two categories—for example, spiders and flowers. Subjects must respond to these stimuli using a joystick, with which they can execute "pull" and "push" responses which cause "zoom" effects (i.e., the pulled/pushed stimulus becomes larger/smaller mimicking actual approach/avoidance). Usually, one of the stimulus categories is expected to automatically evoke either approach or avoidance, so that trials on the AAT can be classified as congruent or incongruent: on congruent trials the instructed movement is the same as the automatic response, and on incongruent trials the instructed movement is opposite to the automatic response. For example, we would expect arachnophobes to tend to avoid stimuli depicting spiders (Rinck & Becker, 2007). The AAT can thus provide measures of performance decrements due to incongruence to measure automatic approach and avoidance tendencies for one stimulus category versus the other. Using the AAT, it has been shown that heavy drinkers with a risk gene for alcoholism are faster at pulling than pushing alcohol cues, as well as other appetitive cues (Wiers, Rinck, Dictus, & van Den Wildenberg, 2009). A conceptually similar task is the Stimulus Response Compatibility task. In this task, a manikin is moved towards or away from a centrally located cue: thus, incongruence can be measured as in the AAT, except with approach/avoidance involving distance between the manikin and the stimulus instead of the zoom-in/zoom-out effect of the AAT. The Stimulus Response Compatibility task has also shown alcohol-approach biases in heavy drinkers (Field, Caren, Fernie, & De Houwer, 2011; Field, Kiernan, Eastwood, & Child, 2008). There is currently much debate on the meaning, interrelationships, replicability of published results, appropriate calculations (correction for control conditions), and optimal design of such implicit measures (Field et al., 2011). Nevertheless, such results indicate that incentive salience indeed plays a role in addiction, although the precise nature of incentive salience remains unclear: Should such
effects be described in terms of stimulus attributes, or in terms of the incentive value of the outcome of an act that can be performed on the stimulus?

The link between incentive salience and physiological states has not as yet been extensively studied in humans, but stimuli with incentive salience would be expected to evoke “hot” physiological states due to their link to an original unconditioned stimulus representing an evolutionarily relevant event. The combination of fundamental biological responses and incentive salience provides a basic model for the abstract metaphor of “heat” to describe psychological or neural processes. Decomposing the idea of “heat” by grounding it in basic biological ANS processes and incentive salience raises the possibility that perhaps effects of “heat” on, e.g., response selection or decision making are effects of these general emotional processes on general selection or decision-making systems, rather than such systems having separate “hot” and “cold” variants. Imagine, for instance, that there is a single response-selection system, consisting of an interdependent set of processes such as encoding and predicting the value of outcomes, the activation of potentially relevant response options, and the outcome-based selection amongst them (Knutson & Wimmer, 2007; Seger, 2008; Wickens, Budd, Hyland, & Arbuthnott, 2007). This system could well be affected by stimuli with more versus less incentive salience. Further, such effects could well be nonlinear: Perhaps qualitatively different behaviors would arise when tasks require responses that are congruent versus incongruent with stimulus-goal associations (see, e.g., work on Pavlovian versus instrumental conflict, and how this may map onto the affect versus deliberation distinction: Dayan, Niv, Seymour, & Daw, 2006). If a unitary system would indeed operate like this, the resulting patterns of behavior could lead to the incorrect assumption that this duality in behavior implies two different systems.
Dual-process and dual-systems models

Models that explain behavior via the outcome of two qualitatively different and competing types of processes or systems are widespread in many fields of psychological theory (Evans, 2008; Schneider & Shiffrin, 1977; Strack & Deutsch, 2004). These processes have been described using various terms (Evans, 2008): As discussed above, the terms “hot” and “cold” are in widespread use, but many other terms have been suggested also, for example, impulsive versus reflective (Bechara, 2005; Hofmann, Friese, & Wiers, 2008; Strack & Deutsch, 2004; Wiers et al., 2007), reflexive versus reflective (Lengfelder & Gollwitzer, 2001), the X- versus C- system (Lieberman, 2007), system I versus system II (Kahneman, 2003), top-down versus bottom-up (Posner & Petersen, 1990), and automatic versus controlled (Satpute & Lieberman, 2006; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Volman, Roelofs, Koch, Verhagen, & Toni, 2011). While important differences exist between these variants, the proposed dichotomies do share a family resemblance. The processes are characterized as less versus more aware, intentional, efficient, and controllable (Bargh, 1994; Moors & De Houwer, 2006), or unconscious versus conscious, implicit versus explicit, low versus high effort, parallel versus sequential (Evans, 2008). Evidence for dual processes comes from a wide variety of studies showing qualitative differences between automatized and untrained performance and uncontrollable effects of manipulations and distractions, implying the existence of automatic processes that may interfere with the controlled processes serving task goals.

There appear to be at least two broad and common types of dual-process models (a third type of multiple-process model is Fuzzy Trace Theory (Reyna & Brainerd, 1995), which differentiates two representational processes—gist vs. verbatim processes—and a third that is related to emotion and inhibition; see, e.g., Rivers et al., 2008 and Chick & Reyna, 2011). First there is a
general information-processing viewpoint, with roots in cognitive psychology, exemplified by the seminal work of Schneider and Shiffrin (1977). They defined automatic versus controlled processing in the context of a model in which memory is conceived of a network of extremely abstract "nodes," which represent any unit of elements related to information processing (e.g., associative connections, response programs, and directions for the processing of information by other nodes). The set of activated nodes is described as being "in working memory," although notably this does not necessarily refer to a separate system in these models. This working memory has a limited capacity; only a subset of nodes can be active at once. An automatic process is a sequence of activation of nodes that occurs in response to an initial input configuration of activation and that progresses without needing control or attention. A controlled process is a sequence of activation of nodes that requires attention, or, that is dependent on "processing-directives" nodes that need to be active in working memory. A classic series of experiments on automatic versus controlled processes from this perspective compared the effects of automatization in visual search tasks (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977): subjects learn to automatically detect stimuli after extensive training. The Stroop task provides another example of automatization: the automatic process of reading words leads to performance deficits when having to name the color in which the word is printed, if the word itself is a conflicting color word (Stroop, 1935). Simon tasks (Simon & Rudell, 1967) and Flanker tasks (Eriksen & Eriksen, 1974) provide further examples of flexible but slow and vulnerable controlled processes, versus reliable and fast but rigid and hence possibly task-inappropriate automatic processes. A difference between these latter tasks and the Stroop task and Schneider and Shiffrin's search tasks is that the involved automatic processes are due to inherent properties of the human attentional system rather than a learning process.
Notably, the above conception of dual-process research does not focus on emotion or motivation (or other "hot" processes); the automatization process is described in highly abstract terms and the tasks involve stimuli and responses that appear highly unlikely to evoke defensive or appetitive physiological responses. In a second line of dual-process research, automatic (or impulsive) processes are far more closely related to emotion and motivation (Strack & Deutsch, 2004). The model underlying such research is closely related to the “horse and rider” metaphor (Hofmann, Friese, & Strack, 2009): Our emotional “animal” drives pull us to immediate reward and away from imminent punishment without regard for long-term consequences, and our rational self must control them and steer us towards virtuous—or, in terms of the lab setting, task-relevant—behavior. As already discussed, the broad class of approach-avoidance tasks shows performance deficits when subjects have to avoid an attractive stimulus, such as a drug cue, or approach an aversive stimulus such as an angry face (Volman et al., 2011), or a spider in an individual with arachnophobia (Rinck & Becker, 2007). Emotional Stroop tasks (Frings, Englert, Wentura, & Bermeitinger, 2010; Williams, Mathews, & MacLeod, 1996) provide evidence for task-irrelevant processes that cause distraction when subjects are exposed to emotional words of which they are instructed to name the color. A task that has been very extensively used to study automatic processes related to evaluation is the Implicit Association Test, or IAT (Greenwald, McGhee, & Schwartz, 1998; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). This is a classification task: subjects must press one or another button to indicate to which of two possible categories a presented stimulus belong. An IAT typically involves two classification pairs: an evaluative classification, e.g., “good” versus “bad”, and a target classification, e.g., “spider” versus “flower.” In the essential part of the IAT, subjects have to perform both classifications in one block. However, they can only use two buttons, so that one response represents one of the target classes as well as one of the evaluative classes, and the other button represents the other
target class as well as the other evaluative class. This results in congruence versus incongruence between target words and evaluative words assigned to the same response button. Thus, in the classic example, spiders and insects can be shown to be automatically evaluated more negatively than flowers because of increased errors and slower reaction times in the incongruent (insects–positive on one response; flowers-negative on the other) blocks than in the congruent (insects-negative; flowers-positive) blocks. Such congruence effects have been used to study automatically activated associations and attitudes involving race, food, politicians, and so forth. We briefly note that it is debated whether IAT scores purely reflect underlying evaluative associations, or whether they may also be due to, e.g., which categories are more salient, or to the selection of exemplars of stimulus categories (Blanton & Jaccard, 2006; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Fiedler, Messner, & Bluemke, 2006; Olson & Fazio, 2004).

The dual-process models discussed above, in particular the emotion-based models, have been applied to theories of addiction, in which the paradox of persistent behavior against the person's own interests and explicit desires has been described as an inability of reflective processes to sufficiently modulate the effects of impulsive processes (Bechara, 2005; Deutsch & Strack, 2006; Stacy, Ames, & Knowlton, 2004; Wiers et al., 2007). The studies discussed above in the context of incentive salience in addiction can often be interpreted in dual-process terms: biases are due to automatic processes evoked by the drug-related stimuli, which lead to impulsive responses or task attentional shifts that conflict with explicit task goals. In line with the incentive salience account, alcohol cues appear to be relatively easy to condition: consistently selecting an alcohol stimulus in a forced-choice task leads to strong automatization in more heavily drinking subjects, as reflected by performance costs when subjects are instructed not to select that stimulus.
(Gladwin & Wiers, 2011a). A sufficiently effective reflective system would be needed to minimize the effects of this conflict between conditioning and task goals, for instance via top-down reduction of the salience of drug cues (Finn, 2002). Of particular interest from a dual-process perspective is that alcohol cues may actively interfere with controlled processing (Gladwin & Wiers, 2011b), potentially leading to a vicious cycle in combination with incentive salience. Dual-process models have received more specific support from findings showing that higher working memory capacity (Grenard et al., 2008; Thush et al., 2008) and interference control capacity (Houben & Wiers, 2009; Wiers, Beckers, Houben, & Hofmann, 2009) weaken the impact of automatic processes on behavior. It appears to be necessary to have both strong associations and weak executive control to show drug-related biases.

Another research domain in which dual-process models—with a strong focus on neural processes—are currently influential, is adolescent behavior, particularly when explaining changes in risky and other possibly problematic behaviors that occur during the transitions from childhood to adolescence and from adolescence to adulthood. From real-world statistics, it is known that adolescents and young adults, compared to both children and older adults, show increased levels of risk-taking behaviors in the form of risky driving, unsafe sex, criminal behavior, and experimentation with and initiation of substance use (Reyna & Farley, 2006). The respective dual-process models of adolescent risk taking share many commonalities with models to explain substance use and addiction, as we have explicitly discussed in Gladwin, Figner, Wiers, and Crone (2011). These frontostriatal neurodevelopmental models of adolescent decision-making posit a potential for an imbalance between strong motivational-affective bottom-up processes and relatively weak controlling top-down processes (Blakemore & Robbins, 2012; Crone & Dahl, 2012; Richards, Plate, & Ernst, 2013; Somerville, Jones, & Casey, 2010; Steinberg, 2010). The
assumption of a developmentally transient imbalance during adolescence is grounded both in animal work (Spear, 2011) and insights from human neuroanatomical development (Giedd, 2008). However, it is important to state that the current empirical evidence that investigated risk taking in children, adolescents, and adults behaviorally and/or neurally is both generally sparse and not unequivocal in supporting or refuting this "imbalance" model: First, the inverted-U developmental trajectory in risk-taking levels across the relevant age range—both observed in realworld statistics and predicted by the imbalance model—appears to be elusive in controlled laboratory situations, as only very few studies observed such a trajectory (Burnett, Bault, Coricelli, & Blakemore, 2010; Figner, Mackinlay, Wilkening, & Weber, 2009b). Consistent with this more anecdotal observation, such a trajectory was also not observed in a formal meta-analysis of the existing risky decision-making studies in the relevant age range (Defoe, Dubas, Figner, & van Aken, submitted). Second, the (still very few) fMRI studies investigating the neural age differences predicted by the imbalance model in both subcortical and cortical neural responses do not provide consistent results: Some studies find evidence consistent with the model (e.g., striatal "hyperreactivity" to rewards in adolescents: Chein, Albert, O’Brien, Uckert, & Steinberg, 2011; Cohen et al., 2010), but others report an absence of age differences or patterns opposite of what the model would predict (e.g., no age differences or striatal "hyporeactivity" to rewards in adolescents: Bjork et al., 2004; Bjork, Smith, Chen, & Hommer, 2010; Paulsen, Carter, Platt, Huettel, & Brannon, 2011; see, e.g., also Pfeifer & Allen, 2012 and Reyna et al., 2011; and for an early behavioral study, see Reyna & Ellis, 1994). Finally, in Reyna and Farley (2006), risky behaviors were not predicted by impulsivity, but by explicit ratings of risk and benefit. It is thus currently difficult to draw any firm conclusions on the imbalance model of adolescence.
Despite the success and/or popularity of dual-process models illustrated in these two illustrative domains of addiction and adolescence—and there are of course many more domains in which such models are widely accepted—we will argue here that the step from the observation of "dualistic" patterns in behavior or brain activation to an underlying model involving dual processes or systems is hazardous. This is especially true for the second kind of emotion-based dual-process models, which explicitly attempt to incorporate motivation and emotion in their "systems," unlike the more abstract information processing models (although a criticism of these latter models could be that they are so abstract that they leave much to be explained, thereby evading the problems the models criticized in this section at least attempt to address). First, the fact that task-irrelevant processes can influence performance does not mean that there exist consistent sets of task-relevant versus task-irrelevant processes. The most careful interpretation of the evidence appears to be simply that cognition is not immune to task-irrelevant effects. Individuals' differences in how badly they are affected are similarly not necessarily due to their having one system that is strong relative to another system; perhaps, for instance, some kind of automatic process that inhibits distracting information leads to high scores on executive control tasks as well as weak effects of task-irrelevant processes. Second, the characteristics that define automatic versus controlled processes do not consistently respond to manipulations as a unit, as would be expected if they reflect a common type of process or the function of one system. In contrast, for instance, a process may be efficient—a property of automatic processes—but still may be dependent on volition and intentions—a property of controlled processes (Bargh, 1994; Evans, 2008; Moors & De Houwer, 2006). An example given by Bargh is driving: Although many components of driving are automatic, drivers do not automatically start driving when seated at the wheel, and where they drive is dependent on where they want to go. This suggests that a simple binary division of the processes underlying behavior into automatic or controlled is
untenable. One response to this is to claim that no task or task manipulation aimed at detecting automatic or controlled processes is process-pure (Conrey et al., 2005): All behavior depends on some mixture of controlled and automatic processes; or, every mental process can have some—perhaps varying—attributes of automaticity and control. However, such a degree of nuances strongly diminishes the parsimony and falsifiability of dual-process models. That is, models with too many “moving parts” provide less and less advantage over simply considering the features attributed to one or the other type of process by themselves, without clustering them into coherent constellations. Third, models positing dual (cognitive or neural) systems have been shown to have far weaker evidence than often assumed (Keren & Schul, 2009; Pfeifer & Allen, 2012). For instance, the finding of brain activation differentially related to one or the other type of processing cannot logically be taken as evidence for separable processing systems. In the elegant model of the X- and C-neural systems (Lieberman, 2007), referring roughly to reflexive and reflective processing, regions that were initially attributed to the X-system, such as the basal ganglia, are for instance also involved in executive control—a function of the C-system (Frank, Loughry, & O’Reilly, 2001; Hazy, Frank, & O’Reilly, 2007; Persson, Larsson, & Reuter-Lorenz, 2013; Van Hecke et al., 2010). Such falsifiability, at least of the details of the model, was predicted by Lieberman and is in principle a scientific strength of the X- and C-model—a good model needs to be falsifiable, and there is clearly room for refinements that could use new information. But the model must also allow a more rigorous falsification, namely of the adequacy of the basic division into dual systems. For example, the “X-system” could be reformulated to consist of networks that implement the fast detection of salience and defensive reactions, such as the amygdala and periaqueductal gray (Hermans, Henckens, Roelofs, & Fernández, 2012), or regions that re-establish stored information-processing procedures, such as the hippocampus and, of potentially central importance to this question, the cerebellum (Marvel & Desmond, 2010).
However, it is unclear from this perspective what the C-system would be, and whether it would actually still form a “dual system”—rather than an opposing system, it would consist of processes that fundamentally rely on the input from the X-system to function at all. Finally, the attribution of motivation and emotion to only one of the systems or sets of processes leads to what has been termed the motivational homunculus problem: That is, when controlled processing is required to “do the right thing” given a certain task, context, or set of long-term contingencies, it must be explained why the control exerted by the subject should be task-appropriate or have a long-term positive expected outcome. Evidently, motivation and control must be interwoven, as opposed to functioning as competing processes. Indeed, there has been increasing interest in the integration of motivation and reinforcement on the one hand and controlled processing on the other (Gladwin, Figner, Crone, & Wiers, 2011; Hazy, Frank, & O’Reilly, 2006; Kouneiher, Charron, & Koechlin, 2009; Pessoa, 2009; Robbins, 2007). However, this again blurs the lines between dual systems.

**The future: asking better questions**
The models discussed above describe interesting and important phenomena, but suffer from concerning flaws. The general underlying problem appears to be premature abstraction: “Hotness” is an intuitively appealing abstraction from physiological states and mesolimbic functions, but studies appear to use the term only as a vague abstraction, without resolving the metaphor to concrete, precise relationships. Similarly, positing the existence of “systems” suggests that we have some knowledge of what these systems are, what they consist of, what set of equations describes them, etc. In contrast, again, they appear to be used more as suggestive placeholders, appealing to common sense and intuition. Due to premature abstraction, studies will be aimed at answering badly defined questions, will be unable to specify precise measures and
operationalizations, and hence will be unlikely to converge on a clear theory. This may play a role in the general methodological problems of psychological research that have recently come under scrutiny. If we don’t really know what we’re looking for, we are far more likely to commit some form of data-snooping or method-snooping—and have far more freedom to do so—to at least find something.

Of course, many researchers have recognized these problems and attempted to deal with them. One question that has been raised concerning dual-process models is whether there are, perhaps, a different number of systems—one, or three, or more. While this skepticism concerning the duality of models is commendable, perhaps we should be questioning the use of “systems” itself. Is defining a system for this and a system for that, a system with these features and a system with those, the best way to understand decision making, response selection, emotion, etc? We briefly note some general approaches that may lead to important alternative ways forward. First, frameworks of dual-process theories could be built more rigorously on conditioning processes (de Wit & Dickinson, 2009; Dickinson & Balleine, 2011), which may provide insight into conflict between well-defined processes related to types of conditioning, for instance by re-interpreting affective versus deliberative processes as Pavlovian-instrumental interactions (Dayan et al., 2006). Second, computational modeling of controlled and automatic processes forces researchers to at least making explicit what we do and do not know and what theories say in exact terms; such models have for instance made clear how working memory may be related to reinforcement (Hazy et al., 2006). Third, we may need the creative generation of novel fundamental types of processes such as iterative reprocessing (Cunningham, Zelazo, Packer, & Van Bavel, 2007), as explored in the next section.
**R3: The Reprocessing and Reinforcement model of Reflectivity**

We have previously suggested a broad class of model, termed the *Reprocessing and Reinforcement model of Reflectivity* or R3 model (Figure 1), in an attempt to address criticisms of dual-process models (Gladwin et al., 2011; Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013). The core of the model is a cyclical process of response selection based on prior reinforcement, in particular of act-outcome associations (de Wit & Dickinson, 2009; Dickinson & Balleine, 2011). In the basic "step" of the model, current-state representations activate a set of associated responses, which activate associated outcomes given the state, which in turn are used to select responses. "Responses" in the model are highly abstract, in the tradition of Schneider & Shiffrin's (1977) nodes, and include all kinds of behavioral and cognitive responses. A response node could include functions traditionally termed executive or controlled, such as an attentional shift, search for information, or update of information in working memory.

We emphasize the time-dependence of activation in the selection process due to iterative reprocessing (Cunningham et al., 2007). The available set of responses may change over time passed since stimulus presentation; and the incentive value of the predicted outcome of an act may also change over time, and is assumed to also be dependent on the state. As an example showing the potential relevance of iterative reprocessing and time dependence, recall that the attentional bias towards and away from alcohol cues in alcohol-dependent patients was highly time-dependent, moving from an initial approach to a—possibly more reflective—avoidance bias (Noël et al., 2006). Control, or reflectivity, is defined within the model as the effective time the response selection process is given to converge by the parameters of the system. That is, reflective processing means that sufficient time is given to allow time-dependent processes affecting response options and incentive values to converge on a stable optimum, which would not be replaced if more processing cycles were completed. Impulsive behavior occurs when
responses are executed after a "too short" reprocessing time, although in certain situations external constraints would make reflective processing disadvantageous.

Figure 1. Illustration of the R3 model

The basic function of the unified system of the R3 model is the selection of the response (R_i) associated with the optimal outcome (O_i), given the current state (S). We show a simple case with only two activated responses. The essential feature of the model is that the activation of responses and the value of the associated outcomes varies over time. This implies that different responses could be selected at different time points following an event. A delay parameter (D) allows the selection process time to settle on non-impulsive responses that require longer to win the competition, represented here via simple lateral inhibition (LI). In the figure, arrows show (often bidirectional) effects on interaction between nodes; arrowheads represent, roughly, activating effects while blunt line-endings represent inhibitory effects. In this abstraction, the output of outcome nodes is assumed to be signed: positive for reinforcing outcomes and negative for punishing outcomes. Delay here is shown to be dependent on the State and the amount of conflict, encoded in the activation of the Lateral Inhibition node. Thus, simply changing the delay parameter can shift processing from a more impulsive to a more reflective state, without needing to assume separate systems. Each node may involve various brain networks. Reflectivity, however, is defined as an emergent property of the system as a whole, and is not assigned to any element within the system. Note further that delay is itself a response with an expected outcome (O_D), that must be associated to appropriate situations via reinforcement.
The R3 model, beyond attempting to capture the nature of reflective or controlled processing, serves as a thought-experiment intended to make four main points. First, the model addresses the motivational homunculus problem: Assuming a purely cold, unemotional control system, the question arises why such a system would serve our interests, why it would be aimed at achieving positive outcomes, even if at longer delays. Our reflective behavior has to be motivated by emotion and incentive as much as our impulsive behavior. The R3 model shows a system in which emotion and motivation, in the form of the incentive value of outcomes, are part of every response selection or decision. “Hot” stimuli could well disrupt the reprocessing cycles within the model, but note that these disruptions take place in the same system that would serve response selection in a “cold” situation with little immediate arousal (but nevertheless always some value to giving the correct response). Second, we used the model to illustrate category mistakes that can be made by ignoring what we term levels of emergence. Features that are relevant to behavioral patterns or subjective thought should not be attributed to elements of a model containing underlying processes. That is, no process is either reflective or reflexive, or controlled or automatic: The system as a whole functions in a more or less reflective state, depending on certain parameters—in particular the time allowed for reprocessing (note that this essential delaying is a response itself, subject to selection and learning processes; it is therefore not a "delaying homunculus," and could well be tuned incorrectly). One could say that every sub-process is "automatic," as reflectivity or control is only defined, or only exists, at the level of the system as a whole. Third, the time allotment serves as an illustration of a single parameter that determines how reflective response selection will be. That is, the model shows that we can have reflective and impulsive states of processing without any kind of conflict between a reflective and
impulsive system, and without having to distinguish subsets of types of processes that can be classified as either automatic or controlled. Possibly interesting from this perspective is that no single brain area, such as dorsolateral prefrontal cortex, which is strongly related to working memory, should contain "control." Recent brain stimulation work provides a possibly interesting illustration of this point. Transcranial Direct Current Stimulation (tDCS) of dorsolateral prefrontal cortex has been found to enhance performance in working memory tasks (Fregni et al., 2005; Ohn et al., 2008); we recently replicated and extended this basic finding using a Sternberg task involving distraction (Gladwin, Uyl, Fregni, & Wiers, 2012). However, the same stimulation protocol failed to reduce the congruence bias on an insects-and-flowers Implicit Association Test (Gladwin, den Uyl, & Wiers, 2012). Even more surprisingly, DLPFC stimulation actually selectively enhanced performance on congruent trials, on which there was no conflict involving evaluation but there was the need to apply a stimulus-response rule. Although this line of research is only starting, such studies start to show that the function of the stimulated region is not "control," but a specific subprocess that could either aid or hinder the efficiency of reflective processing. Finally, the model changes the focus from an unspecific “strength” metaphor to a learning perspective that emphasizes how successful reflective processing at least partly results from an individual’s reinforcement history. Has an individual been rewarded for responding to certain situations with, e.g., delaying responses, or with a slow memory search strategy to look for downsides to immediate response options? Does an addict, beyond being burdened with fast and easily available drug-seeking responses associated with high incentive value, have any reinforced alternatives to which response selection could converge given sufficient time? The model appears to fit very well with the “Tools of the Mind” approach (Diamond, Barnett, Thomas, & Munro, 2007) for reinforcing the use of executive function in young children, and is perfectly illustrated by the "Think About The Answer Don't Tell Me" song (which can be found
by searching for "Diamond - Day Night Presentation.mpg" on the internet). We briefly mention that the model also extends the "binding" concept—the need to temporarily associate elements in memory—from stimulus features and stimulus-responses connections to act-outcome associations. This is beyond the scope of this chapter, but may connect dual-process models to psychophysiological methods and results involving phase coding and synchrony (Gladwin, 't Hart, & de Jong, 2008; Gladwin, Lindsen, & de Jong, 2006; Jensen & Lisman, 1998).

To summarize, we first aimed to clarify terminology by describing the various uses of temperature metaphors, and attempted to decompose "hotness" into basic biological responses and incentive salience. We then extended the scope to discuss dual-process and dual-system models. These models are highly successful (surely also in the sense of being popular), but nevertheless need to be viewed with scientific skepticism, which we discussed next. Importantly, changes to the models and subsequent lines of research may have implications for a wide variety of applications, such as cognitive bias modification protocols in addiction and other psychopathology (Wiers et al., 2013), in our understanding of adolescent behavior (Pfeifer & Allen, 2012), or in theory- and evidence-based approaches in education (Diamond et al., 2007). Finally, we briefly presented some possible ways forward, including our own R3 model.

Regardless of the routes researchers will choose, we believe that the time is ripe for models to more often become objects of critical study rather than accepted assumptions.
References


Van Hecke, J., Gladwin, T. E., Coremans, J., Destoop, M., Hulstijn, W., & Sabbe, B. (2010). Prefrontal, parietal and basal activation associated with the reordering of a two-element list
held in working memory. *Biological psychology*, 85(1), 143–8. doi:10.1016/j.biopsycho.2010.06.005


