European Journal of Personality
Eur. J. Pers. **24**: 376–403 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/per.779

Automatic and Controlled Processes in Behavioural Control: Implications for Personality Psychology

PHILIP J. CORR*
University of East Anglia, UK

Abstract
This paper highlights a number of unresolved theoretical issues that, it is argued, continue to impede the construction of a viable model of behavioural control in personality psychology. It is contended that, in order to integrate motivation, emotion, cognition and conscious experience within a coherent framework, two major issues need to be recognised: (a) the relationship between automatic (reflexive) and controlled (reflective) processing and (b) the lateness of controlled processing (including the generation of conscious awareness)—phenomenally, such processing seems to 'control' behaviour, but experimentally it can be shown to postdate the behaviour it represents. The implications of these two major issues are outlined, centred on the need to integrate theoretical perspectives within personality psychology, as well as the greater unification of personality psychology with general psychology. A model of behavioural control is sketched, formulated around the concept of the behavioural inhibition system (BIS), which accounts for: (a) why certain stimuli are extracted for controlled processing (i.e. those that are not 'going to plan', as detected by an error mechanism) and (b) the function of controlled processing (including conscious awareness) in terms of adjusting the cybernetic weights of automatic processes (which are always in control of immediate behaviour) which, then, influence future automatically controlled behaviour. The relevance of this model is illustrated in relation to a number of topics in personality psychology, as well related issues of free-will and difficult-to-control behaviours. Copyright © 2010 John Wiley & Sons, Ltd.

Key words: automatic processing; behavioural control; behavioural inhibition system; consciousness; controlled processing; personality

INTRODUCTION

Mechanisms of behavioural control (e.g. automatic vs. controlled processing) are fundamental in psychological explanation; and individual differences in these mechanisms may be assumed to play an equally important role in personality psychology. As Carver,

*Correspondence to: Philip J. Corr, University of East Anglia, UK. E-mail: p.corr@uea.ac.uk

Johnson, and Joormann (2008) note, studies from cognitive, social, personality and development psychology converge on the conclusion that there exist (at least) two modes of information processing and action regulation, which operate simultaneously and often in competition with each other. The interplay of these mechanisms may be especially important in accounting for individual variation in both normal and abnormal behaviour.

This paper contends that psychology in general, and personality psychology in particular, has failed to grasp firmly enough the implications of multiple levels of behavioural control and, in consequence, theoretical advance and integration are being impeded. The aim of this paper is to highlight some of the problems that need to be recognised and addressed, and to invite commentary on a proposal for their resolution.

Consideration of the importance of multiple levels of behavioural control requires the recognition of a number of key theoretical issues; namely (a) the relationship between automatic (reflexive-non-conscious) and controlled (reflective, often with conscious representation) processes and (b) the lateness of controlled processes (including the generation of conscious awareness)—phenomenally, such processes seem to 'control' behaviour, but experimentally they can be shown to postdate the behaviour they represent. This paper is primarily concerned with behavioural control; however, issues surrounding consciousness and related phenomena cannot be ignored, especially because so many concepts in personality psychology assume (sometimes explicitly, but more often tacitly) the involvement of consciousness. It is contended that, as a result of a failure to come to terms with the implications of the above theoretical problems, personality psychology continues to be characterised by a plethora of different theories, each tending to be focused on a single level of control and presented in a way that renders their integration with other theories, at best, difficult. This state of affairs was noted by Corr and Matthews (2009, p. xxxviii–xxxix) in their Introduction to the *Cambridge Handbook of Personality Psychology*,

> A persistent theme. . .has been the multi-layered nature of personality, expressed in individual differences in neural functioning, in cognition and information-processing, and in social relationships. Abnormal personality too is expressed at multiple levels. Despite the inevitable difficulties, a major task for future research is to develop models of personality that integrate these different processes.

To illustrate the importance of the above problems for personality psychology, they are discussed in relation to two topics: (a) affect and emotion and (b) personality measurement. A general model of behavioural control is proposed, based on experimental and clinical neuropsychological data, and formulated around the well-established *behavioural inhibition system* (BIS; Gray, 1982; Gray & McNaughton, 2000; McNaughton & Corr, 2004, 2008a).

The relevance and importance of the BIS for the arguments and the model presented in this paper are underscored by the fact that the BIS forms the basis of an explanation of how automatically processed information (i.e. pre-potent behaviour) gets extracted and subjected to higher-level cognitive analysis by controlled processes. Not only does the BIS provide an explanation for this transition, in the form of a mismatch error signal (i.e. between expected and actual states of the world), but Gray's (2004) functional model of consciousness, which is an extension of BIS theory and upon this the proposed model is based, offers an explanation for conscious awareness: It seems to afford the facility to inhibit pre-potent responses when they are inappropriate to the demands of the environment. This causal (i.e. mismatch error signal) aspect of BIS theory, coupled with

DOI: 10.1002/per

the inhibitory functions of conscious awareness, justifies the role assigned to BIS theory in this paper. This choice is not to discount the value of other theoretical approaches, but few of these theories address the above 'transition' and 'functional' problems; and none, to my knowledge, within a broader consideration of the 'lateness' of controlled processes (see below). In addition, BIS theory is well established in personality psychology and already contains many of the elements required for the construction of a general model of behavioural control.

It is to be hoped that the general form of the proposed model will help to provide theoretical guide-ropes for the construction of more specific models of behavioural control in personality psychology. It is acknowledged that the proposals in this paper are merely a prolegomenon to this ultimate goal.

THE PERSONALITY PROBLEM TO BE ADDRESSED

Do personality psychologists need to be concerned with mechanisms of behavioural control, let alone the accompanying exotica of consciousness? Indeed, is there a problem that needs addressing at all? As argued below, there are, indeed, real theoretical issues, and these principally are centred around how controlled (often, but not necessarily, conscious) processes interface with automatic (pre/non-conscious) processes.

As is widely known, controlled processing is not synonymous with conscious awareness; however, the latter often accompanies the former. The precise role played by conscious awareness in controlled processing is open to considerable debate; but to the extent that conscious awareness, and the controlled processes that underlie it, play *any* role, and in order for them to have causal efficacy, they *must* interface with the machinery that controls immediate behaviour which, as discussed below, is controlled at a pre-conscious, automatic level.[1] Therefore, to the extent that conscious awareness, and its underlying processes, form any part of a personality theory, the issue of behavioural control becomes crucial as do the theoretical problems entailed by the lateness of these controlled processing. (There can be few personality theories in which the controlled processes related to conscious awareness are of *no* relevance?)

LEVELS OF BEHAVIOURAL CONTROL

This section presents some of the 'scene-setting' material in preparation for the articulation of the major theoretical problems. It should prove useful in avoiding any misunderstandings concerning the nature of the model proposed.

Cognition

It is important to be clear as to what is meant by 'cognition', especially in the way it differs from 'non-cognitive' (e.g. 'biological') explanations. The concept of 'cognition', as used

[1]This paper is not concerned with the nature of consciousness *per se*. It is concerned with how processing at the controlled level, which often has *representation* in conscious awareness, relates to processing at the automatic level. In a closed physical-causal system, the *mental* aspect of conscious awareness (the *experience* of it) should be clearly differentiated from the mechanisms that control it, the latter of which interfaces with automatic levels of control.

DOI: 10.1002/per

in this paper, refers to the capacity to know and to have knowledge; and this definition includes the structures and information processes that support knowing/knowledge. This knowledge and the processes of 'knowing' are embedded in structures, beliefs and operations (e.g. decision-making) that, in a fundamental conceptual sense, exist independently of nervous activity (although, of course, they are instantiated in this activity). For example, *knowledge* of Renaissance art, as contrasted with Cubism, is not determined by nerve assemblies—although, it should not be forgotten that our visual perception of art is determined by nervous system activity (e.g. the construction of the qualia[2] of colour from electromagnetic reflections from the paint surface). This knowledge is *often*, but need not be, accessible to conscious awareness; however, to avoid the ever-present Cartesian trap, it is *not* assumed that conscious awareness comprises or controls the underlying cognitive mechanisms; rather, it is seen as one of outputs of controlled processing.[3]

Thus, one major problem that *any* theory of cognition and behaviour must address—*to the extent that cognition is different from motor control processes*—is how knowledge-level structures/processes interface with biological structures/processes of the neuroendocrine system to affect *immediate* behaviour. In cybernetic terms, cognitive knowledge structures/processes must interface with behavioural systems in order to set the weights at critical points in the regulatory feedback system that choreographs and controls behaviour—as elaborated below, behaviour is *always* initiated and executed at a pre-conscious, automatic level: Mind events follow brain events. This is a basic tenet of materialist brain science, which in one form or another is the standard model endorsed (or, at least, not openly disavowed) by (the majority of) contemporary researchers.

Preconscious and non-conscious processing

There is now considerable evidence to show that there are whole classes of events that never reach conscious awareness. Velmans (1991) reviewed a large experimental literature from which he concluded that all of the following processes are capable of being, and normally are, completed pre-consciously—that is, before there is any conscious awareness of what has been carried out: (a) analysis of sensory input; (b) analysis of emotion content and input; (c) phonological and semantic analysis of heard speech; (d) phonological and semantic analysis of one's own spoken words and sentences; (e) learning; (f) formation of memories; and (g) choice and preparation of voluntary acts. The phenomenon of 'blindsight' (i.e. subjective blindness but intact visual performance; Weiskrantz, 1986) is a specially telling example of how actions can be controlled by non-conscious (automatic) processes. In addition, experimental evidence shows that some events that are usually accessible to conscious awareness can also be processed non-consciously (e.g. reactions to emotion-inducing faces below the threshold for awareness; e.g. Feng, Luo, Liao, Wang, Gan, & Luo, 2009). The ubiquity of non-conscious processing is consistent with the

[2]Qualia (singular is 'quale') is a term used in philosophy to denote the subjective quality of mind, referring to the way things seem to us (from the Latin 'what sort' or 'what kind') in the form of properties of sensory experience such as sensations (e.g. pain) and percepts (e.g. colour).
[3]The existence of conscious awareness remains a mystery, as most of its putative functions may, in principle, be completed successfully at the automatic level of processing (see Velmans, 1991), and there are a number of possible explanations for it, none of which are theoretically satisfactory or scientifically compelling (see Corr, 2006).

DOI: 10.1002/per

(re)discovery of implicit personality processes (Bargh & Williams, 2006; for further evidence, see below).

Dual-process models

The need to differentiate levels of behavioural control is demonstrated by the wide variety of dual-process models in the literature (e.g. Carver, 2005; Eisenberg, 2002; Epstein, 1973, 1994; Evans, 2003; Hirsh, 1974; Lieberman, Gaunt, Gilbert, & Trope, 2002; Metcalfe & Mischel, 1999; Rolls, 1999; Rothbart & Bates, 2006; Rothbart, Sheese, & Conradt, 2009; Strack & Deutsch, 2004; Toates, 1998, 2006; see Carver et al., 2008). Most of these models contain a combination of the following features:

1. *Automatic (reflexive)*: Fast, coarse-grained, ballistic (implicit/procedural learning), and pre/non-conscious.
2. *Controlled (reflective)*: Slow, fine-grained, deliberative (explicit/declarative learning), and often accessible to conscious awareness.

The necessity of assuming, at least, two relatively autonomous systems suggests that evolution had to negotiate two major conflicting demands; that is, how to achieve adaptive 'fast and dirty' behavioural responses, especially in defensive reactions, as well as 'slow and clean' behavioural responses, especially in complex or novel environments (LeDoux, 1994).

A good illustration of these processes in personality psychology is contained in the model of Ortony, Norman, and Revelle (2005) who postulated three levels of control: Reactive, routine, reflective, each with affect (feelings), motivation (needs/wants), cognition (knowledge, thought and beliefs) and behaviour (action). The reactive and routine levels are comparable to an automatic, reflexive system, while the reflective level is comparable to slower and more deliberate controlled forms of cognitive processing.

A major theme of this paper is the need to unify theories *within* personality psychology, providing a conceptual bridge for the two-way flow of theoretical traffic. Understanding the respective functions of automatic-reflexive and controlled-reflective processes is necessary as is their interaction, including the crucial question as to why some information is extracted and subjected to fine-grained analysis. These issues are not new. Cognitive and social psychology have for many years addressed these problems (e.g. Schneider & Shiffrin, 1977), although personality psychology has tended to remain devoted to theories that focus, to a significant extent, on outputs of controlled processing available to conscious introspection (e.g. self-concepts). There are notable exceptions to this trend in the areas of, for example, congruence between implicit and explicit motives and emotional well-being (Langens, 2007); implicit motives and sexual motivation and behaviour (Schultheiss, Dargel, & Rohde, 2003); implicit motives and self-attributions (McClelland, Koestner, & Weinberger, 1989); implicit self-esteem (Farnham, Greenwald, & Banaji, 1999); dissociations between implicit and explicit personality self-concepts (Asendorpf, Banse, & Mücke, 2002; Hofmann, Gschwendner, Friese, Wiers, & Schmitt, 2008); role played by implicit personality (Greenwald, Poehlman, Uhlmann, & Banaji, 2009); and different levels of control in anxiety (Ouimet, Gawronski, & Dozois (2009).[4]

[4]I am grateful to an anonymous reviewer for highlighting the importance of this literature.

DOI: 10.1002/per

THE ARROW OF CAUSATION

The literature summarised above draws attention to a fundamental issue that has not been sufficiently acknowledged in personality psychology; namely, the fact that controlled processing comes *after* the relevant brain-behavioural event (indeed, some 300–500 milliseconds)—that is, on a millisecond-by-millisecond basis, the engagement of controlled processes and their *re*presentation in conscious awareness lags behind the brain's initiation and execution of the behaviour itself. In consequence, only the *results* of the processes are accessible to conscious awareness, not how the behaviour was initiated and executed (the production of language is an obvious example of this distinction). This is the 'lateness' of controlled processes, which has been the target of considerable empirical attention, as discussed in detail below.

One immediate objection to the above statement is that 'cognition' and 'controlled processes' can operate at a relatively automatic level and, therefore, are not 'late' in the causal chain of events. This may be true; however, it is not true of *all* cognition and controlled processes: These are the ones that present the 'lateness' problem. Turning to consciousness, the cognitive processes that control conscious awareness come before the time required for the generation of conscious awareness, so it may seem that the problem of lateness is merely a problem for consciousness that we can then dismiss as having no causal role to play in behaviour. However, these controlling cognitive processes are also late in the causal chain of events. There are good reasons to assume—or at least, not to pre-maturely dismiss—the possibility that the controlled processes involved in conscious awareness do have important causal functions; and, to the extent that they do have such functions, they pose a 'lateness' problem.[5] To the person in the street, and even to professional psychologists (save radical behaviourists), to suggest otherwise would invite a near-consensus of derision.

In everyday life, as well as in many personality theories, it is the very content of the *re*presentation contained in controlled processing and consciousness that is of central importance. Clearly, self-belief, meta-cognition, etc. are, to some extent, accessible to conscious awareness, and they are certainly related to high-level, complex cognitive processes (Robinson & Sedikides, 2009). This contention is consistent with the notion that the self is hierarchically organised, with its most abstract features captured when individuals characterise *themselves* (Schell, Klein, & Babey, 1996). If this were not the case, then self-report would be impossible. Self-concepts, beliefs, etc., that *are* to a large extent consciously mediated, are the theoretical substance of much of personality psychology. They are assigned causal roles—or, at least, their underlying cognitive processes are assigned such a role—and given prominence in many theories of personality. However, there is the conundrum of how they achieve any degree of causality, given that they (both awareness and their underlying cognitive processes) seem to come too late in the causal chain of events to which they refer. Thus, there is an important problem to solve.

[5]One route around this problem would be to assume that conscious awareness and its related processes are causally impotent, and that all behavioural control takes place in an automatic manner. However, this would encounter the major problem of explaining (away) all forms of consciousness and would, in consequence, strip-bare concepts of self-concept, free-will, etc. from psychology. However, Gray (2004) provides a cogent reason, relating to its systematic nature, for accepting that the processes underlying conscious awareness have an evolutionary and, thus, functional, role in behavioural control.

DOI: 10.1002/per

LATENESS OF CONSCIOUS AWARENESS

Since the 1950s, Libet (1985; for a summary, see Libet, 2004) has conducted a series of ingenuous experiments to show that it takes upwards of ~500 milliseconds of brain activity for conscious awareness to be generated: This is the 'lateness' of conscious experience—by inference, we can extend this lateness to all high-level controlled processes. In one series of studies, the somatosensory cortex of awake patients was directly stimulated with trains of pulses—such stimulation leads to sensory perception (e.g. of being touched) (Libet, 1982). What was intriguing about these studies was the finding that there appeared to be a necessary period of 'neuronal adequacy', involving ~300–500 milliseconds of continuous stimulation, before consciousness was experienced. Various control experiments confirm the robustness of this observation.

Such findings pose a problem for any adaptive theory of consciousness, and related controlled processing, because long before ~300–500 milliseconds, motor actions have already been initiated. For example, the removal of the hand from a hot stove occurs before awareness of the hand touching the stove. However, an intriguing twist to these findings is that events are not experienced as if they happened ~300–500 milliseconds ago: Consciousness appears to refer to what is happening *now*. Libet proposed the provocative suggestion that the conscious experience of a stimulus is 'referred back in time', tagged to the readiness potential (RP) that occurs close to the beginning of the initiating brain action.

Concerning the volition of will, Libet went on to explore absolute timing using conscious intentions. Briefly, the typical experiment required participants to note the instant they experienced the wish to perform a 'voluntary' action (e.g. simple flexion of finger)—that is, the instant they were consciously aware of the wish to act. To record this time, participants remembered the position of a revolving spot on a cathode ray oscilloscope, which swept the periphery of a face like the second hand of a clock (but much faster so as to achieve more sensitive time estimates). During this time, the RP from the motor cortex was recorded by EEG. This procedure allowed Libet to calculate the precise moment at which the participant 'decided' to make the movement, and then to compare the timing of this moment with the timing of events in their brains. He found evidence that these 'conscious decisions' lagged between ~350 and ~400 milliseconds behind the onset of the RP—once again, the conscious wish came a long time *after* the brain started to initiate the action, but subjectively it did not feel that way to experimental participants. Debate of Libet's findings (e.g. Libet, 2003; Zhu, 2003) has not undermined the robustness of the basic phenomenon of the lateness of conscious awareness. Similar effects can be demonstrated in experimental (non-human) animals, where recording of cell activity in the hippocampus may be used to predict resulting behaviour (e.g. Ferbinteanu & Shapiro, 2003; Morris & Hagan, 1983; Wood, Dudchenko, Robitsek, & Eichenbaum, 2000).

Well, Libet's findings may well be interesting and intriguing, but are they of relevance, let alone importance, for personality psychology? In terms of cognitive consciousness (e.g. knowledge-level beliefs that can be verbally expressed), the answer is clearly affirmative, as this comes hundreds of milliseconds after the automatic processing to which it refers. As discussed throughout this paper, this lateness poses multiple problems for personality constructs that are related to cognitive consciousness, and which are so often endowed with causal properties. (There is no less of a problem if we talked exclusively about non-conscious, high-level, controlled processes as these too lag behind automatic processing.) Faced with this answer, there are two possible positions to adopt: The epiphenomenalist's stance; that is, consciousness is created alongside certain types of brain information

DOI: 10.1002/per

processes, but lies outside the causal chain by which such processing produces behaviour (there are good reasons for rejecting this position; see below); or the position adopted here, that they do play a causal role in the control of behaviour, but not in the manner commonly assumed. Accepting the latter position, the pertinent question is: How do (mental) conscious representations, or more précisely their underlying (cognitive) controlled processes, influence automatically processed behavioural routines? The remainder of this paper is devoted to providing a tentative answer to this question.

SUMMARY OF THE PROBLEM

The problem to be addressed by the model of behavioural control may now be summarised. At the point of initiation and execution, *all* brain-behavioural processes are controlled by the automatic-reflexive system, and the operations of this system cannot be affected simultaneously by high-level controlled processes, and nor can they be consciously known as only their products are *re*presented in conscious awareness. In order to eschew a dualistic position, brains events *must* precede mind events, *always*.[6]

Now, if controlled processing and conscious awareness comes only *after* corresponding brain events and is the outcome, or product, of such causally sufficient processing, then how do controlled-reflective (often, but not necessarily, conscious) processes exert any influence (if they do) on automatic-reflexive (pre/non-conscious) processes? This is a central question in general psychology and personality psychology. It resides at the core of the issue of how multiple level processes interface; and how personality factors and processes operate at and between these levels.

Is there really a problem at all?

Habits of thought and theory in personality psychology militate against the unreserved acceptance of the above assertions. Other researchers have noted a similar reluctance to accede to the causal priority of pre-conscious events. For example, in relation to action (dorsal stream) and perception (ventral stream) visual systems, Goodale and Milner (2006, p. 663) noted,

> The most difficult aspect of our ideas for many people to accept has been the notion that what we are consciously seeing is not what is in direct control of our visually guided actions. The idea seems to fly in the face of common sense. After all our actions are themselves (usually) voluntary, apparently under the direct control of the will; and the will seems intuitively to be governed by what we consciously experience. So when we claimed that a visual illusion of object size (the Ebbinghaus illusion) did not deceive the hand when people reached out to pick up objects that appeared to be larger or smaller that they really were, vision scientists around the world embarked on a series of experiments to prove that this could not possibly be true.

[6]This statement may seem unnecessarily proscriptive—even dogmatic. And it may be tempting to speculate that there may exist some form of controlled (executive functions) processing supervisory control system that governs automatic processes (including the exercise of free-will). Alas, the invocation of such a control system does not negate the veracity of this statement, because such a system would be subject to exactly the same problems that it attempts to circumnavigate: it would come too late in the causal chain to be able to acquire any causal control on immediate behaviour. What 'supervisory system' exists can only exist in the form of automatic routines.

DOI: 10.1002/per

FUNCTIONS OF CONSCIOUS AWARENESS

In order to lay down the foundations of the model of behavioural control outlined in the next section, this section considers the possible functions of controlled processing and conscious awareness based on the work of Gray (2004) who addressed the problem of the functions of consciousness (and its related cognitive processes) from a philosophically sensitive, neuropsychological perspective. It does not attempt to offer an account of the 'Hard Problem' (Chalmers, 1995); that is, the *why* and *how* of conscious experience, especially how the brain *generates* conscious awareness. Instead, it focuses on the *functions* of consciousness: What is it *for* and *how* these functions are implemented? One way of identifying these functions is to ask: In what ways do the occurrence of conscious awareness, and related processes, lead to behaviour that is different from, and has greater survival value than, behaviour that does not entail such awareness? Obvious answers to this question are refuted by the fact of conscious awareness coming late in the causal chain of brain-mind events.

The assumption that consciousness *per se*, and its underlying processes, do have a survival value, and fitness-enhancing functions, is suggested by: (a) inter-individual consistency (as far as we know)—a lack of such a value would tend to lead to genetic drift and less obvious consistency; (b) qualia that allow the classification and differentiation of evolutionarily significant environmental stimuli (e.g. nutritious vs. poisonous foods) and (c) behavioural evidence to suggest that non-human animals too have something similar to human consciousness (this conclusion is supported by the considerable psychological similarities observed across the phylogenetic scale; see McNaughton & Corr, 2008b).

Gray's model addresses the observation that the events of which we become conscious are neither a random nor complete set of those events of which we could, in principle, become conscious, given that the brain receives information about them. Unlike other models, Gray takes seriously the implications of Libet's delay of conscious awareness and in doing so posits three linked functions of consciousness.

(1) It contains a model of the relatively enduring features of the external world; and the model is experienced as though it *is* the external world.[7]
(2) Features that are particularly relevant to ongoing motor programs, or which depart from expectation, are monitored and emphasised.
(3) The control variables and set-points of the brain's non-conscious servomechanisms are juxtaposed, combined and modified; in this way, error can be corrected.

Late error detection

According to Gray's (2004) model, the principal function of the conscious processing system, and its related cognitive processes, is the interruption of automatic brain-behaviour routines ('reflexes') that are not 'going to plan' (i.e. where an error signal has been detected). Central to this model is a 'comparator', which compares actual stimuli with expected stimuli—this function is performed by the BIS (Gray, 1982; Gray & McNaughton, 2000; McNaughton & Corr, 2004, 2008a; for a summary, see Corr, 2008). When there is no discrepancy, and 'all is going to plan', the comparator is said to be in 'just checking mode' and behavioural routines run uninterrupted and do not experience a transition to controlled processing mode. However,

[7]In addition to Gray's (2004) book, excellent presentations of the subjective construction of the external world are given by Frith (2007) and Ramachandran (2003, 2005).

DOI: 10.1002/per

when a mismatch is detected, between the actual and expected states of the world, then the comparator goes into 'control mode', and the salient features of the error-triggering environment are extracted and subjected to controlled, attentional, analysis and (often) *re*presented in conscious awareness (the contents of consciousness are constructed qualia that categorise and *re*present information from the external world; e.g. colour categories from the continuous spectrum of, colourless, electromagnetic energy).

Inhibition of pre-potent behaviour

Automatic routines are well suited to reacting to predictable stimuli from a pre-existing behavioural repertoire; however, such automatic behaviours are not so good for tasks requiring a departure from fixed routines (e.g. a novel task), or when automatic behaviour is not going to plan. Much of cognitive processing involves inhibitory functions, and the 'late error detection mechanism', activated when things are not going to plan, serves this function well.

An experimental demonstration of the power of conscious awareness to inhibit pre-potent (automatic) responses is seen in the 'Jacoby exclusion task' (Debner & Jacoby, 1994). Briefly, words are presented either too fast for conscious recognition (i.e. 50 milliseconds) or slow enough for recognition (i.e. 150 milliseconds); backward masking is used to ensure these precise presentation times. In this experimental paradigm, participants are presented with the prime-word, for example:

H O U S E

They are then given a stem-completion task, for example:

H O U _ _

A possible stem completion is to add S and E to form 'HOUSE'.

Now, the crucial manipulation in this task is the instruction to participants *not* to complete the word-stem with a prime-word. In the above example, it might be completed with N and D to form 'HOUND'.

This task is trivially easy for most people, but *only* when the word is presented above the threshold of awareness (at 150 milliseconds). What happens when the prime-word is presented *below* the threshold of consciousness? In this case, there is an inability to follow the instruction not to complete the word-stem with the presented prime-word. In fact, what happens is that the word-stem is completed *more often* with the covertly presented prime-word, HOUSE rather than HOUND (or some other word completion). It, thus, seems that the default reaction to a word-prime presented covertly is to prime the word-stem, and that the generation of conscious awareness is needed to prevent this automatic priming effect—the fact that the conscious mind can prevent this priming effect demonstrates its power to inhibit pre-potent automatic reactions. This result seems to point to something important about conscious awareness: *Somehow*, the generation of conscious awareness, and its underlying cognitive processes, enables the inhibition of pre-potent (automatic) responses.

This inhibitory mechanism solves one major evolutionary problem: How to ensure that automatic brain-behavioural routines are appropriate. It would be desirable to have the ability to inhibit the firing-off of these automatic routines in some circumstances (e.g. inhibiting avoidance behaviours when in foraging mode), even if this inhibition takes several hundreds of milliseconds, which is usually enough time to have important consequences on behaviour.

To illustrate these inhibitory functions, imagine a person is confronted by a dangerous snake. Their fear system would be activated and automatic brain-behavioural routines (e.g.

DOI: 10.1002/per

simple fleeing reaction) triggered. This activation would be initiated long (i.e. hundreds of milliseconds) before the person was aware consciously (i.e. 'see' and 'feel' qualia) of the event. Now, it would be highly adaptive for them to have the facility to 'replay' this immediate past in order to analyse its contents—this facility would be especially beneficial at times when their automatic behavioural routines did not achieve their goal (e.g. avoiding the dangerous snake in the first place).

The general form of Gray's model is consistent with other models of action awareness. For example, Farrar, Franck, Paillard, and Jeannerod (2003, p. 618) noted,

> This function is achieved through a comparison process between the predicted sensory consequences of the action, which are stored in its internal model, and the actual sensory consequences of the action.

Gray's (2004) innovation was to suggest that there is an error-triggering mechanism to the generation of conscious awareness, linked to the BIS. The notion that the functions of consciousness are tied closely to cognitive control, especially behavioural inhibition, is commonplace in consciousness studies. For example, Dehaene et al. (2003) noted that the anterior cingulate (ACC) is active during a variety of cognitive tasks that entail mental effort, and that its involvement in these tasks may be explained by its role in the detection of conflicting response tendencies, although only when conflicting stimuli are consciously perceived.

In a similar vein, Mayr (2004, p. 145) observed,

> It has been often noted that we usually become aware of those aspects in the internal or external world that interfere or interrupt routine action—which are very same events that typically elicit executive control operations.

Thus, BIS activation by error-triggering stimuli may be seen as the precursor to executive functions that then control subsequent information processing. As the BIS is seen by many researchers as one of the fundamental systems underlying human defensive behaviour and negative emotionality (for a review, see Corr, 2008), this BIS-related error-triggering mechanism may be seen to be of prime importance in personality psychology. Arguably, it offers a coherent theory to advance future research.

Individual differences in the sensitivity of the BIS should be expected to determine the threshold for this error-triggering mechanism, the inhibition of pre-potent behaviour, and the generation of the contents of consciousness. At the high pole of the BIS dimension (as, for example, measured by the Carver & White, 1994, BIS/BAS scales), we should find highly anxious individuals whose BIS is in a chronic state of over-activation, entailing worry, rumination and risk assessment (cognitive qualia), behavioural inhibition and high levels of negative emotionality. At the low pole of the BIS dimension, we should find individuals with an under-active BIS, entailing an impaired ability to detect goal-conflict (i.e. mismatch between expected and actual stimuli), a lack of inhibition of inappropriate pre-potent behaviour, and a general absence of behavioural inhibition, worry/rumination and negative affect. High BIS activation resembles clinical anxiety; and some theorists have suggested that low BIS activation contributes to the emotional, motivational and behavioural characteristics of primary psychopath (e.g. Lykken, 1995; for a BIS-based neuropsychology model of psychopathy, see Corr, 2010).

Gray's (2004) theory is compatible with other models of consciousness, for example Baars' (1997) theory of global workspace. Upon the workspace 'blackboard' of Baars'

DOI: 10.1002/per

model, error-prone information is written and subjected to further processing. According to Baars, consciousness is similar to a bright spot on the theatre stage of Working Memory (WM), directed by a spotlight of attention under executive guidance (Baddeley, 1986). Continuing with this metaphor, the rest of the theatre is dark and unconscious. With Baars' model, working memory is important because it has the function of disseminating information to various modules throughout the brain. Gray's theory proposes *why* information is extracted and subjected to the spot-light of working memory and cognitive processing that leads, often, to conscious experience.

In conclusion of this section, the inhibitory function of consciousness solves one major evolutionary problem: How to ensure that *proximal* automatic responses are appropriately activated; and how *distal* controlled processes are invoked only at critical junctures, when a definite choice has to be made and a cautious, risk-assessing, mode of processing is more appropriate than the pre-potent response. At these critical junctures, and after fine-grained analysis afforded by controlled processing, cybernetic adjustments can be made to the automatic system, such that when the same (or similar) stimulus (e.g. snake) is encountered in the future, automatic-reflexive behaviour will be more appropriate. This process of behavioural adjustment can happen over the course of hundreds of millisecond, which can, and do, result in life-or-death outcomes (e.g. feedforward planning in predator–prey struggles). In this way, *distal* (controlled-conscious) effects come to influence (automatic/non-conscious) *proximal* effects, albeit with a time lag.

DEFENSIVE SYSTEMS OF BEHAVIOUR

The above discussion of the functions of consciousness has taken place in relation to the BIS, which is part of the reinforcement sensitivity theory (RST) of personality (Corr & McNaughton, 2008; Gray & McNaughton, 2000; McNaughton & Corr, 2004, 2008a) which comprises two other major systems, discussed below. RST provides a convenient model of the automatic processes involved in approach and avoidance behaviour with which to start to build a model of behaviour control.[8]

In brief, RST comprises three systems as follows:

(1) The *fight–flight–freeze system* (FFFS) is responsible for mediating reactions to *all* aversive stimuli, conditioned and unconditioned, and is responsible for avoidance and escape behaviours. It mediates the emotion of fear, and the associated personality factor consists of fear-proneness, timidity and avoidance.
(2) The *behavioral approach system* (BAS) mediates reactions to *all* appetitive stimuli, conditioned and unconditioned, and is responsible for approach to appetitive stimuli. It mediates the emotion of hope and anticipatory pleasure, and the associated personality factor consists of optimism, reward-orientation and impulsiveness.
(3) The BIS is responsible for the detection and resolution of goal-conflict in very general terms (e.g. between BAS-approach and FFFS-avoidance), and evolved to permit an animal to withhold entrance (i.e. passive avoidance) or to enter a dangerous situation (i.e. leading to cautious 'risk assessment' behaviour), such as a foraging field where predators may be hiding. Its principal function is to resolve the evolutionarily

[8]The systems of RST are not exclusively automatic-reflexive as they have representations at all levels of the behavioural hierarchy. However, they embody many such features, especially at the lower and more primitive levels of defensive reactions.

DOI: 10.1002/per

important conflict resulting from risk-aversion (FFFS) and risk-proneness (BAS). The BIS is involved in the processes that finally generate the emotion of anxiety.

In broader psychological terms, the BIS entails the inhibition of pre-potent, conflicting, behaviours, the engagement of risk assessment processes, including the scanning of memory and the environment to help resolve concurrent goal-conflict. This is all experienced subjectively as worry, apprehension and the feeling that actions may lead to a bad outcome. The BIS resolves goal-conflicts by increasing, through recursive loops, the negative valence of stimuli (held in cortical stores), via activation of the FFFS, until resolution finally occurs either in favour of approach or avoidance. Theta activity is the neural signature of this BIS activity and can be identified by theta EEG coherence during emotionally charged rumination (for a discussion and empirical confirmation, see Andersen, Moore, Venables, & Corr, 2009).

Figure 1 shows the flow of information in automatic and control modes. Information is extracted and subjected to fine-grained analysis by higher-level controlled (cognitive) processes (which sometimes, but not always, is accompanied by conscious awareness) under conditions where an error signal is generated; that is, under conflict between the expected and actual state of the world (which includes the inner world of planning and

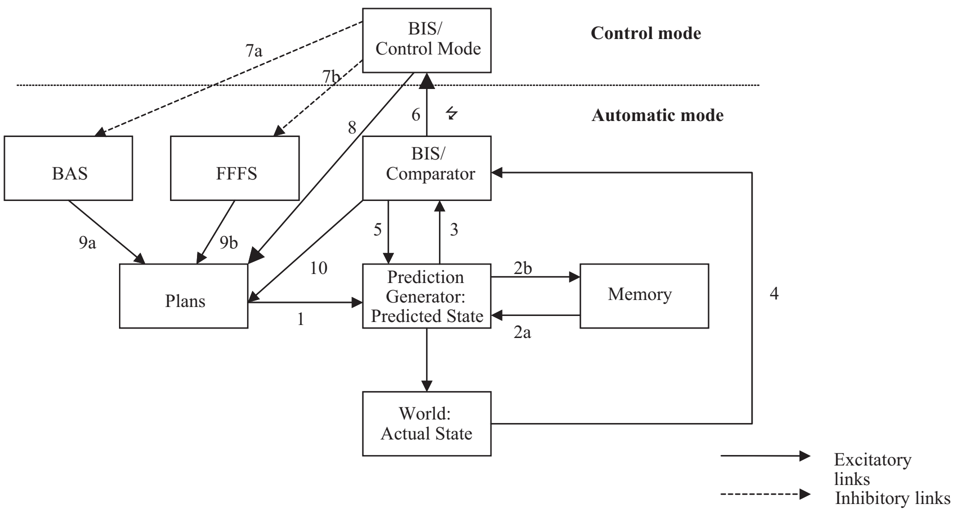

Figure 1. The conceptual functioning of the behavioural inhibition system (BIS) in ‘automatic mode’ and ‘control mode’ processing, containing basic approach (BAS) and avoidance (FFFS) factors that, along with the BIS, comprise three fundamental systems and dimensions of personality. Behavioural plans (Plans) lead to predictions (Prediction Generator; 1) of future states of the world, which receives input from (2a), and sends output to (2b), stored previous experience (Memory). The BIS (BIS/Comparator) receives input from the Prediction Generator (3), and then compares the response-reinforcement outcomes (World: Actual State) with predictions (4), and then one of two things happen: (a) ‘everything is going to plan’, and the BIS/Comparator sends input to the Prediction Generator to continue the motor program (‘just checking mode’; 5); or (b) the BIS/Generator detects a mismatch between prediction and outcome and generates an error signal (ϟ), which leads to activation of controlled processes (BIS/Control Mode; 6). Once the BIS/Control Mode is activated, there is inhibition of the behavioural approach system (BAS; 7a) and the fight–flight–freeze system (FFFS; 7b); and at this time the BIS initiates cautious approach and risk assessment (see text), which then informs Plans (8), which simultaneously receives input, about current states, from the BAS and FFFS (9a, b), as well as input, about the nature of the conflict, from the BIS/Comparator (10). Plans initiate appropriate behaviour and the above cycle is repeated, until behavioural resolution is achieved in the form of FFFS-mediated avoidance/escape or BAS-mediated approach.

DOI: 10.1002/per

priorisation of goals). Individual differences in the parameters of this model determine the sensitivity of the system.

These FFFS/BAS/BIS processes are well established in personality psychology and allow specific predictions, concerning the relationship between personality factors and behaviour, to be specified. For example, an over-active BIS will lead to an error signal at a lower level of goal-conflict; the sensitivity of the BAS and FFFS will have a significant impact upon goal planning, and when they are sufficiently and (approximately) equally activated they will trigger goal-conflict by causing the planning and predictor systems to fail to decide upon a dominant form of behaviour: This leads to behavioural dithering, cognitive indecision and control processing. The problem finally resolves itself by the whole system becoming more risk averse and when conflict continues, in consequence, behavioural control reverts to FFFS-mediated avoidance/escape. If the error signal was a false alarm, then behaviour reverts to previous behaviour.

Important in this sequence of events are coping and appraisal mechanisms (both primary and secondary) which have a significant impact on predictions (including self-efficacy) about the world. Therefore, individual differences in sensitivity and activation of FFFS, BAS and BIS give rise to the personality components of this model, as do personality concepts related to self-efficacy, perceived control and appraisal of the consequences of mismatch between the expected and actual state of the world.

FFFS and BIS behaviours are arranged according to a hierarchical system of defence, distributed across multiple brain systems that mediate specific defensive behaviours associated with level of threat experienced (i.e. 'defensive direction'), ranging from the pre-frontal cortex (PFC), at the highest level, to the periaqueductal gray, at the lowest level.

A higher-level, cognitive, virtual reality environment (that is experienced in the medium of conscious awareness), in which the world could be modelled in order to run 'what-if' simulations, may be assumed to have conferred enormous evolutionary advantage in situations where automatic processes were insufficient, for example in complex social situations where conspecific 'politics' was important (as seen today in chimpanzees; De Wall, 2000), although the advantages accruing to this high-level simulation would also be seen in much less complex contexts (e.g. predator avoidance and disgust associated with dangerous foods).

Neurobiology of the BIS

According to BIS theory, the following systems are implicated in the detection of goal-conflict and inhibition of pre-potent behaviour. Detection of simple goal-conflict is based in the hippocampus as the main locus; however, it can be detected at all levels of the BIS, ranging from the periaqueductal gray, medial hypothalamus, amygdala, septo-hippocampal system, and posterior cingulate to the pre-frontal dorsal stream. Lower levels of the defensive hierarchy are responsible for detecting conflict between quick-and-dirty goal representations and produce simple fast responses (such as defensive quiescence); whereas activation of higher levels of the defensive hierarchy are responsible for detecting conflict between slow-and-sophisticated goal representations. In terms of the attentional processing entailed by controlled processes, neurotransmitter systems, principally, acetylcholine and norepinephrine, are likely to be heavily involved (Gray & McNaughton, 2000). Behavioural inhibition, to some extent, is controlled by the inferior frontal gyrus, or under very tight time constraints the pre-supplementary motor cortex (Aron, Durston, Eagle, Logan, Stinear, & Stuphorn, 2007; Aron, Fletcher, Bullmore,

DOI: 10.1002/per

Sahakian, & Robbins, 2003; Floden & Stuss, 2006). In terms of the inhibition of pre-potent behaviour, inhibition involves output from the BIS to whatever motor areas provided the input that generated the conflict. The output will be to the lower levels of the motor system, leaving the activation of the goal representation itself intact but preventing its normal capture of the motor system. Lastly, activation of the BAS, FFFS and the BIS is likely to lead to high levels of arousal, especially emotional arousal via the amygdala, which serves to invigorate behaviour.

SPECIFICATION OF BEHAVIOURAL CONTROL MODEL

The general model of behavioural control can now be outlined. It is an updated model of Corr (2006) and is shown in Figure 2. According to Toates (1998), a stimulus has a given strength of tendency to produce a response; that is, a stimulus has a response-eliciting potential, which varies from zero to some maximum value (this strength depends upon innate factors and learning). According to Gray (2004), a mismatch between what is expected and what is experienced (i.e. the ‘error signal’; either in stimuli or behaviour, including response-reinforcement contingencies) leads to the salient stimuli being extracted and subjected to controlled process analysis and (often) displayed in a medium that is experienced as conscious awareness.

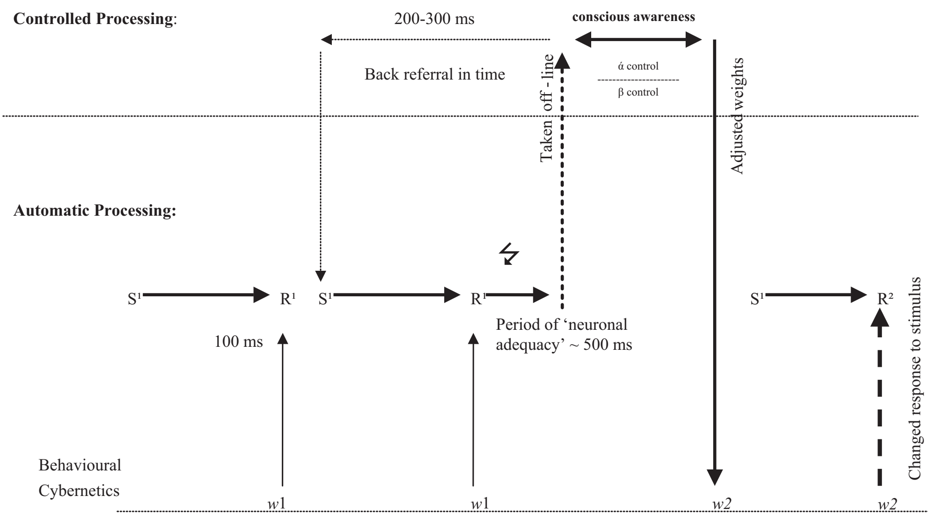

Figure 2. Late error detection model of the functions of consciousness. When ‘everything is going to plan’, automatic processes proceed uninterrupted. When an error signal (ϟ) is detected (i.e. mismatch between expected and actual state of the world), the salient stimuli features of internal (e.g. memory) and external worlds are extracted and subjected to detailed controlled processing, which *may* result in a perceptual-based *re*presentation and display in a medium that is experienced as conscious awareness (‘ά control’ mode); however, controlled processing may occur in the absence of conscious awareness (‘β control’ mode). In either form of processing, extracted stimuli are subject to (varying degrees) of fine-grained analysis—all of this happens within hundreds of milliseconds. Although conscious experience (and related cognitive processes) lag behind automatic processes, crucially, controlled processing can alter the cybernetic weights (e.g. *w*2) of automatic processes and, thereby exert a causal influence on future automatically controlled responses (e.g. R^2) when the same (or similar) stimuli/worlds are encountered. Subjectively, this process is seamless, and importantly, the lag in causal effect is obscured by ‘back referral in time’, which provides the illusion that the experience is occurring at the same moment as the stimuli that it represents.

DOI: 10.1002/per

The proposed model adopts Gray's idea that actions that are organised at the automatic-reflexive level (e.g. fleeing from a predator) can, nonetheless, be affected by controlled-reflective processes.[9] The proposed model extends this idea; for example, a fear state that is experienced consciously has the capacity to sensitise the whole defensive system (e.g. in a particular environment) and, thereby, affect all *subsequent* fast, automatic responses in that environment; alternatively, more specific controlled process alterations can be made to automatic behavioural routines. In other words, over the course of hundreds of milliseconds, behaviour is modified by experience: *Learning* occurs. Thus, controlled process analysis exerts a *distal* influence on automatic routines by *general* and *specific* influences, that is by the changing of specific automatic cybernetic weights such that when the same stimuli, that previously led to an error signal, is encountered again a different (more appropriate) reaction occurs. Controlled and automatic processes differ specifically: (a) in their temporal characteristics; (b) their level of analysis; and (c) their representation in conscious awareness.

Gray's model is predicated on Libet's lateness of consciousness awareness, and the proposed model adopts this idea too. The proposed model, however, proposes a distinction between 'α control mode' (entailing conscious awareness) and 'β control mode' (not entailing conscious awareness). In the 'α control mode' it is the underlying cognitive processes that interface with automatic processes. There is still much to learn about the benefits conferred by the *experience* of conscious awareness, although inhibition of automatically controlled pre-potent responses seems to be one of these benefits.

Executive control

A high level of coordination is needed to ensure flexible behaviour, involving attention, decision-making and integrative functions. Whilst the hippocampus (and other distributed structures) of the BIS may be necessary to mediate error signals, they work in conjunctions with cortical stores of information reflecting the conflicts between *goals*. In addition, activation of the PFC is also expected to be important. With complex behaviour that entails even a modicum of conflict, there is potential for behavioural interference. PFC has been assigned an important role in resolving this behavioural problem. Miller and Cohen (2001) provide a review, and an outline of a model, of how the PFC functions to achieve this coordination. They note that, in order to avoid this behavioural confusion, mechanisms must have evolved that coordinate low-level sensory and motor processes according to the representation of internal goals—this view fits snugly with the cybernetic view of behavioural control advanced in this paper, as well as with the view of the BIS as a goal-confliction detection/resolution device.

PFC is a network of neocortical areas that send/receive projections from nearly all sensory and motor systems, and many subcortical areas. The 'top-down' functions of PFC are guided by internal goal states; and these are especially important when there is a mapping between sensory input, cognition and action that are either weakly developed,

[9]In terms of a computer analogy, controlled processing may be likened to a high-level programming language (e.g. FORTRAN), and automatic processing to machine code of 0s and 1s which is the code recognised by the computer hardware (i.e. the central processing unit; CPU). In order for any such code (be it FORTRAN, propositional logic, imagery, or one's local spoken language) to gain any influence on the machinery that controls behaviour, it must be 'compiled' into the 'language' of the operating system (in the case of neural system, neural networks that interface with sensori-motor systems). A computer compiler performs such operations as: lexical analysis, pre-processing, parsing, semantic analysis, code generation, and code optimisation; and the human neuronal system must do something similar: this takes time and resources and, in addition to the time taken to extract and analyse information, gives rise to the lateness problem.

DOI: 10.1002/per

relative to existing ones, or are rapidly developing. As noted by Miller and Cohen (2001), one of the most important aspects of cognitive control is the ability to select a weaker, task-relevant response, in the face of stronger competing, pre-potent, but task-irrelevant, ones. In this regard, PFC is seen to be important in executive control, especially the active maintenance of goals and the rules of the task required for coordinated behaviour. These authors note that the Stroop task provides a good illustration of this competition, as does the Wisconsin Card Sort Task (WCST), which '...are variously described as tapping the cognitive functions of either selective attention, behavioural inhibition, working memory, or rule-based or goal-directed behaviour' (p. 170). Performance on both tasks has been long known to be impaired in frontal-lobe damaged patients.

Selective attention and behavioural inhibition may be seen as two sides of the same coin: Attention is the effect of biasing competition in favour of task-relevant information, and inhibition is the consequence that this has for the irrelevant information. PFC-mediated control is complemented by another form of control dependent on the hippocampal system (Miller & Cohen, 2001). The hippocampus is important for binding together information into a specific episode; in contrast, PFC, like other neocortical areas, is more important for extracting the regularities across episodes (corresponding to goals and task rules). Miller and Cohen (2001) associate the anterior cingulate cortex (ACC) with the detection of goal-conflict, but there are reasons for thinking that it is not exclusively responsible for this function; see Andersen et al., 2009; McNaughton & Corr, 2008a). In terms of the feedforward functions of the proposed model in this paper, PFC may be assigned the important role of mediating learning across time, by the active maintenance of goals and plans, which provide the necessary temporal bridge. The above conclusions are consistent with BIS theory, which contends that the hippocampus, and other components of the defensive system (e.g. ACC), interact with the neocortex where goals and plans are held, and that this interaction can be recorded as BIS-related theta coherence between these regions (Andersen et al., 2009).

Prediction error signal

It is known that reward learning is related to a 'prediction error signal', related to dopamine (DA) projections from the midbrain ventral tegmental area (VTA). VTA DA neurons burst with activity in response to unpredicted, desirable stimuli, such as food (Mirenowicz & Schultz, 1994; see Bayer & Gilmcher (2005)). As this DA-related error prediction signal seems to be restricted to appetitive, and not aversive, stimuli (Mirenowicz & Schultz, 1996), it has been suggested that there may exist a complementary, independent, system related to aversive stimuli predictive errors, linked to serotonin neurons in the brainstem (Daw, Kakade, & Dayan, 2002). Although little is known about the signals encoded by individual serotonin neurons, there is some evidence suggesting that this neurotransmitter plays an important role in the control of behaviour by aversive events, punishment and losses. For example, animals with lesions of the serotonin system have difficulty acquiring stimuli associations that require the inhibition of a response and difficulty inhibiting the response if the lesion is produced following training (Harrison, Everitt, & Robbins, 1999; Soubrie, 1986). In terms of the proposed model, serotonin is seen as one of the most important neurotransmitters as it innervates the entire defensive system (including both FFFS and BIS; see McNaughton & Corr, 2008a). It is noteworthy that the treatment of choice for many neurotic disorders entails enhancement of serotonin neurotransmission (e.g. Selective Serotonin Reuptake Inhibitors; SSRIs). The above views are consistent with

DOI: 10.1002/per

Carver et al. (2008), who claim that serotonin is important in the transition to/from automatic/controlled processing.

Free-will

As an example of how controlled and automatic processes interface in the model, consider the issue of free-will. The model proposed here challenges one version, namely ‘primary free-will’ (i.e. actions follow conscious volition). However, other versions of free-will are less damaged, especially ‘secondary free-will’, which entails the influence of controlled processes on *future* behavioural (automatic) routines (‘future’ here can refer to hundreds of milliseconds). A more immediate form of ‘free-will’ is the *veto* of automatically initiated actions (in Libet’s term, ‘free won’t’); the free-will to inhibit. This inhibitory veto occurs after the initiation of the automatic behaviour but before the full sequence of behaviour is executed. Although time delays are quite long (~300–500 milliseconds for conscious awareness to be generated), over seconds (and minutes or hours), there is considerable causal influence from conscious awareness in terms of (re)setting the cybernetic weights that determine the next automatic action sequence. Added to this would be the cybernetic weights related to ‘free won’t’—some people seem particularly hesitant in their thoughts and behaviour, and these may reflect automatic built-in circuit breakers (inhibitory veto) to automatic behavioural routines. As an example, lying entails a considerable amount of ‘free-won’t to inhibit pre-potent truth and for this reason can often be detected by a high degree of hesitation and cognitive dithering.

A summary of the model

The main points of the foregoing discussion may be summarised as follows:

(1) Many of the variables falling under the rubric of ‘cognition’ (especially high-level controlled processes, those available to conscious awareness, and those involving concepts of the self) come too late in the causal chain of events to affect proximally the behaviour they *re*present.
(2) Cognition need not involve high-level controlled processes or conscious awareness, but then these forms of ‘cognition’ (e.g. priming) do not differ in fundamental respects from automatic-reflexive behaviour (they may still be relatively complex, e.g. language comprehension)—in this way, pre/non-conscious cognition does not pose a problem for the model (but such cognition would need to be stripped of any ‘late’ components).
(3) In relation to point 2, it may be asked:
    (a) To what extent are beliefs, values, intentions, etc. automatic-reflexive and to what extent are they controlled-reflective?
    (b) If such beliefs, values and intentions are, indeed, automatic-reflexive, then how can they differ, in fundamental terms, from similarly automatic ‘biological’ processes (e.g. basic defensive reactions)? At this point of synthesis, ‘biological’ and ‘cognitive’ levels collapse to a single automatic-reflexive level (e.g. as seen in multiple implicit processes in personality psychology); and as such, the only problem remains to show how truly controlled-reflexive processes relate to automatic-reflexive processes.
(4) Behaviour may be modified by controlled-reflective processes by changes to the cybernetic weights of the automatic-reflexive system, thus affording behavioural flexibility, adaptation and learning.

DOI: 10.1002/per

(5) *All* behaviour, *at the moment of initiation and execution*, is the proximal result of automatic-reflexive processes (*ex hypothesi*, they may have been modified by prior controlled-reflective activity).
(6) According to the model, automatic-reflexive and controlled-reflective processes serve very different functions, are compatible, and may be integrated into a unified general theory of behavioural control.
(7) Individual differences in the operating parameters of components of the behavioural control system give rise to individual differences in affect, motivation, cognition and behaviour and are, therefore, of fundamental importance in personality psychology.

Some general implications of the model

Individual differences *within* these two major systems of behavioural control, as well as their *interplay*, should account for important sources of variance between people. Some potential implications are outlined below.

First, a person could have all the 'will' (i.e. high-level controlled processing and conscious desire) in the world to behave in a certain way (e.g. dieting), but their 'will' can only translate into actual behaviour if the controlled processing system is able to interface effectively with the automatic processing system that, in a proximal sense, controls immediate behaviour (e.g. priming effects by hunger). Secondly, difficult-to-stop emotions/behaviours feature prominently in personality psychology (as well as in many psychiatric disorders). In the case of emotional engagement and expression, especially as seen in the dysfunction of regulation in mood disorders, automatic defensive reactions are often difficult to stop or inhibit (e.g. depressive rumination and violent rage)—drugs may directly inhibit these automatic processes, but 'talk therapy' (e.g. cognitive-behavioural therapy) would also have the power to modify the cybernetic weights of these automatic processes by engaging controlled (usually consciously-mediated) processes. Thirdly, there may be insufficient representation by controlled processes in automatic processes, leading to hard-to-stop counter-productive behaviours. For example, cigarette smoking may be difficult to stop because there is more salient (in terms of priming) representation in the automatic-reflexive processes than controlled-reflective ones.

In relation to the last point, it is interesting to consider the issues surrounding substance abuse. Extensive work by Robinson and Berridge (1993; Berridge & Robinson, 1995), on the distinctions between 'wanting' and 'liking', has shown the importance of distinguishing between automatic and controlled processes. Robinson and Berridge (1993, p. 72) states,

> The neural systems that are most sensitized by drugs normally mediate a specific motivational process we call 'wanting' or, more formally, attribution of incentive salience. This psychological process is not 'liking' or pleasure, nor is it directly experienced in conscious awareness. Nonetheless, it causes the perception or representation of an event to become attractive, sought after, and capable of riveting attention. . .people do not have the direct conscious awareness of either wanting or liking. Rather, activation of the neural substrates of wanting or liking must be translated into subjective awareness by cognitive mechanisms, as are other complex perceptions. Because the basic processes that mediate wanting and liking are not directly accessible to consciousness, people may find themselves wanting particular things without knowing why. Under some circumstances, people may not even know that they want them.

In further support of the distinction, it is interesting to note that prolonged conscious introspection distorts judgements of pleasure (Wilson & Schooler, 1991), suggesting that conscious awareness can conceal rather than reveal true affective liking. This issue is discussed further in the 'affect and emotion' section, below.

DOI: 10.1002/per

Some evidence adduced in support of the model

At the outset it should be said that there is a paucity of evidence directly addressing individual differences at the *interface* of automatic and controlled processes, although personality psychology attests to the wealth of evidence—some of which has been reviewed in the paper—relating separately to automatic and controlled processes. The automatic-controlled interface may be illustrated in relation to two studies specifically focused on the clinical condition of depression and the personality trait of neuroticism.

In the clinical field, depression is associated with impairments in (explicit) episodic memory, but spared implicit memory (Jermann, Van der Linden, Adam, Ceschi, & Perroud, 2005). This literature suggests that depressed patients are impaired in their ability to use effortful (conscious) processing (both encoding and retrieval). Using the Process Dissociation Procedure (PDP; Jacoby, Toth, & Yonelinas, 1993)—which enables a distinction to be made between automatic (via familiarity judgments) and controlled (via recollection data) processing—Jermann et al. (2005) studied the effects of inclusion instructions (i.e. complete the word-stem with the prime-word) and exclusion instructions (i.e. do not complete the word-stem with the prime-word). In the exclusion condition, controlled and automatic processes work in opposite directions (non-conscious priming vs. conscious inhibition), creating interference. Results reveal that clinically depressed patients, compared with normal controls, have lower estimates of controlled processing, but their automatic processing is intact. (For further discussion of this process distinction in depression and other forms of psychopathology, see Carver et al., 2008.)

In the personality field, one study serves to illustrate dysfunctions in automatic and controlled systems interface. Corr (2003) showed that individuals who are high scorers on the trait of neuroticism show impaired automatic processing (i.e. the procedural learning of the sequence of stimuli) *in the presence* of controlled (attentional) dual-task processing.

How may the above two studies be related in terms of the behavioural control model proposed in this paper? High neuroticism individuals, with a hyperactive BIS, would tend to experience a high level of error signals, which would lead to material being extracted and subjected to controlled process analysis. Now, in conformity with Jermann et al.'s (2005) finding with depressed patients, if controlled (explicit) processes are defective (possibly because they are being overwhelmed by the cognitive demands imposed by the extraction of such a large quantity of material), then high neuroticism individuals would fail to resolve goal-conflicts effectively[10] and would, thus, show an impaired ability to adjust the cybernetic weights of the automatic system (which is responsible for the pre-potent defensive behaviour, negative affect, etc.)—in computer parlance, the high-level language would be ill-specified and would not be read properly by the compiler which is responsible for the transition from controlled to automatic information. This conclusion is consistent with the association of low serotonin with (a) neuroticism (and related clinical disorders) and (b) disrupted transition to/from automatic-controlled processing (Carver et al. 2008).

Clearly, more focused research is needed in order to clarify the precise nature of the automatic-controlled interface in different clinical conditions and in relation to the major dimensions of personality.

[10]As with many processes in psychology, there is assumed to an optimal level of BIS activity. At very high levels, goal-conflicts would be easily, and inappropriately, triggered and would consume the resources of the system to resolve them effectively—they may not even be solvable goal-conflicts in the first place—and this failure to resolve goal-conflict would, itself, be another form of goal-conflict, setting in train a vicious, and pathological, cycle.

DOI: 10.1002/per

IMPLICATIONS OF MODEL FOR PERSONALITY PSYCHOLOGY

The stage has now been reached where the implications of the model for personality psychology may be explored in a little more detail. Although the model has implications for many areas of personality psychology, for the sake of brevity the following two areas are discussed: (a) affect and emotion and (b) personality measurement.

Affect and emotion

Although for many years considered a symptom of muddle-minded thinking (e.g. Skinner, 1953), the topic of emotion has been in the ascendency in recent years, and in recognition of its importance there is now the APA journal *Emotion*. The problem that the topic of emotion presents is that if the conscious experience of it comes only after the behavioural event to which it refers, then the generation of emotion may, at first blush, seem pointless and without function. What functions could it serve? This question is far from new, for example, William James (1884) famously asked, ‘What is an emotion?’

In conformity with the model proposed in this paper, one function may be to provide automatic-reflexive level valence to context (e.g. foraging field *x* is potentially dangerous; the expected hostility of a particular audience). This valence bias may then be imposed upon incoming pre-conscious information, such that during the ~300–500 milliseconds ‘neuronal adequacy’ period automatic routines are biased to respond in specific ways to general context, as opposed to specific stimuli. This biasing would take the form of primary appraisal. By this route, controlled processing could have a direct causal influence on automatically elicited behaviour in the next iterative cycle of behaviour. (Being in a ‘bad mood’ perhaps provides an everyday illustration of this function.) This form of valence bias would serve the additional function of priming the inhibition system so that inappropriate pre-potent behaviour is halted faster the next time it generates an error signal.

Secondly, emotion may provide the appropriate valence to control processing representations over the longer time frame, especially when ‘what-if’ simulations are computed to determine the likely future outcome of a specific action. The outcome of these simulations would be to affect the cybernetic weights of automatic behavioural routines. As shown in Figure 1, individual differences in the sensitivity of reward and punishment systems may be expected to influence the ease of generation of these emotions, thus allowing systems such as the FFFS and BAS to affect the contents of consciousness (for a summary of relevant studies, see Gomez & Cooper, 2008). We may also suspect that many of the symptoms seen in various clinical conditions consist in the pathological running of these mental ‘what-if’ simulations (e.g. worry).

The model of behavioural control adumbrated above helps to dissolve another of the venerable debates in psychology, namely that engendered by the James-Lange theory of emotion. According to the model, all emotion must be, *at the point of initiation and execution*, non-conscious, but that cognitive processes are also involved and important, in the ways outlined above. Control can, and is, exerted by automatic processes as well as controlled processes, but the *proximal-distal,* respectively, distinction is important here—indeed, the model predicts that, in the case of adequate inputs to emotion systems, both levels of control would be simultaneously activated. This is an example of how the model might help to avoid the all-too-often ‘switching’ (comparable to a Necker Cube switching) in theoretical perspective, allowing both perspectives to be held in mind at the same time.

DOI: 10.1002/per

Another debate that the model may be useful for throwing light upon is the Larazus-Zajonc debate of the 1980s. Whether emotion is primed via a non-conscious route (Zajonc's position), or shaped primarily through cognitive appraisal processes (Larazus's position), has divided researchers, although support exists for both viewpoints. It is possible to reconcile two seemingly opposing theoretical positions by recourse to different levels of behavioural control. This literature has been well documented, and repetition here would not serve any useful purpose.

The above assertions are consistent with the work of Baumeister, Vohs, DeWall and Zhang (2007), who contend that the view that emotion has a direct causation on behaviour is seen to be increasingly untenable. Instead, they argue that emotion is part of a feedback system whose influence on behaviour is typically indirect: By providing feedback and affording retrospective appraisal of actions, conscious emotional states facilitate learning and change future behaviour. These authors review a large body of empirical evidence to justify their conclusions. What the arguments of this paper contribute to this literature is to highlight the lateness of higher-order cognitive and emotional processes. The general tone of the present argument is in good agreement with their statement,

> The assumption that the purpose of full-blown, conscious emotion is to cause behaviour directly appears to be widespread and indeed deeply embedded in psychological theorizing. Yet it appears to be far less true than many researchers (ourselves included) have assumed. (p. 194)

In relation to substance abuse (also see above), Robinson and Berridge (1993, p. 74) make a similar point,

> An important postulate of our hypothesis is that conscious awareness has only indirect access to attributions of incentive salience. It may seem strange to assert that people are not directly aware of their own likes and wants. After all, whether or not people know much about anything else, don't they know what they like?...Regarding addiction, a consequence of the separation of elementary psychological processes from conscious awareness is that it is not non-sensical to speak of non-conscious wanting or of unconscious pleasure, just as it is not non-sensical to speak of implicit knowledge or unconscious perception.

This line of argument is supported by evidence from different fields of enquiry. For example, people's self-reported romantic partner preferences greatly diverge from their actual choices, suggesting that they are indeed unaware of them (Todd, Penke, Fasolo, & Lention, 2007).

Baumeister et al. (2007, p. 195) go on to claim, which is also endorsed by the arguments and evidence of this paper, that 'Evidence indicates that conscious emotion is helpful for learning' and 'Emotion stimulates reflection in prior events'. At present, we simply do not know why *conscious* emotion should serve this role. This is, admittedly, an enormous lacuna in our theoretical knowledge. However, this lacuna should not devalue the contribution of empirical evidence (see above) that points to the involvement of conscious processing in inhibiting inappropriate pre-potent responses which enables reflection and facilitates the production of more appropriate future behaviour.

Personality measurement

Consideration of multiple levels of behavioural control may also hold important implications for how personality is measured. Allport (1927) raised a question that still awaits an answer: What is the basic unit of personality? To this, should be added: How is it to be measured? For example, are different forms of measurement required for consciously mediated meta-cognitive processes and low-level defensive reactions?

DOI: 10.1002/per

Questionnaire measures remain valuable instruments in personality assessment, and are widely used even to measure biological processes (e.g. Carver & White's, 1994, BIS/BAS scales). However, what a person says they would do in a given situation and what they would actually do is not, of necessity, the same thing. Asking people to say how they *typically* respond is one way around this problem; however, it is far from being unproblematic.

Clearly, to read a questionnaire item (e.g. 'I worry a lot') requires a degree of conscious control (although most of the language processing and conceptual skills needed to read and understand this question are non-conscious), but there is no reason to assume that the response does not reflect, to some extent, the product of automatic processing: Perhaps, for this reason, psychometricans often construct questionnaires that require the respondent to give the first answer that comes into their head and not to think too much about the meaning of the question—we have already seen above that prolonged thinking can distort emotion-based judgments. Perhaps questionnaires have proved valuable precisely because they have not relied exclusively upon controlled (conscious) processing. In any event, questionnaires need to be validated against external criteria; for example, the Fear Survey Schedule (Wolpe & Lang, 1977) against behavioural reactions to exposure to snakes, spiders, etc. Another method would be to test the validity of questionnaires against implicit personality tests (IPT; e.g. Asendorpf et al., 2002).

Defining what is to be measured is a prerequisite to deciding how to measure it. The model proposed here suggest that this is a key feature and that, for example, developing taxonomic models of personality based exclusively upon lexical items may lead to a specific model that would not be so easily replicated using items that preferentially tapped a different level of processing. Taking this line of argument seriously, it may be possible to integrate the five-factor model with, for example, more biologically inspired measures of the FFFS, BAS and BIS (e.g. the Carver & White, 1994, scales) (McAdam & Pals, 2006). Critical to this success would be appreciation of the fundamentally different roles, and implications of, the specific functions of different levels of behavioural control. For example, the lexical nature of the five-factor model may preferentially reflect controlled processing and conscious awareness that codifies important qualities of society (e.g. appreciation of artistic beauty; Openness), the importance of social interactions (Agreeableness) and the value placed by society on detail focus and industry (Conscientiousness). In contrast, measures of the FFFS, BIS and BAS may preferentially reflect pre-potent, emotion-biased, responses which, often, defy rational explanation (e.g. fear of harmless spiders), reflected in various measures of Neuroticism. Extraversion is another example of, largely, automatically-elicited preferences: Whether a person prefers to go to lively party or to have a quiet night at home is not subject to rational judgment; they 'just do'—further enquiry would probably serve only to distort this basic preference by adding controlled processing levels of justification and rationalisations.

CONCLUSIONS

This target paper has highlighted a number of, what have been argued to be, fundamental issues raised by consideration of the multiple-level processes in behavioural control; specifically, implications of the lateness of controlled processing (especially, but not exclusively, entailing conscious awareness) and the interface of automatic-reflexive and controlled-reflective processes. It was argued that these issues are important for personality

DOI: 10.1002/per

psychology, especially with respect to the integration of different theoretical perspectives within this field, as well as their integration with wider psychology.

These theoretical issues may be seen in the broader context of the relationship between behavioural functions and neural activity; an otherwise happy (though not untroubled) marriage before the ménage-a-trios intrusion of the qualia of consciousness and its related controlled processes (Gray, 2003). Although it may be tempting to ignore this intrusive interloper, this would provide only a temporary solution and, for reasons outlined, an inadequate one.

It was argued that a lack of recognition of these theoretical issues continue to impede developments within personality psychology. In an attempt to frustrate this trend, and perhaps by so doing offering some tentative solutions, a general model of behavioural control was sketched, formulated around the well-established concept of the BIS. The model proposed that the BIS comprises a late-error detection device that is responsible for extracting stimuli, actions, etc. ('goals' in the broadest sense) that are not 'going to plan' which are then subjected to fine-grained analysis, whilst at the same time inhibiting error-related pre-potent behaviour. Whilst the discussion was focused on the RST of personality, its implications extend to other theories and families of theories. Arguably, RST encompasses some of the important approach, avoidance and conflict mechanisms that many personality psychologists believe underlie temperament/personality (although not always in the precise form assumed by RST).

The general form of the proposed model share similarities with other models of consciousness (e.g. Gazzaniga, 1988); however, it may be argued that, unlike these other models, it reflects the unique 'holy trinity' of behavioural control: *Lateness* (of controlled processes); *inhibition* (of automatic processes) and *behavioural adaptation* (modification of the cybernetic weights that determine future, automatically-executed, behaviour).

The arguments presented in this paper, and the conclusions reached, will not satisfy all (or even, very many) readers. Nevertheless, it is to be hoped that some important theoretical issues have been aired and, perhaps even, some useful solutions offered; but this theoretical proposal is little more than a sketch of the elements of a viable model of behavioural control that still awaits development.

ACKNOWLEDGEMENTS

I am grateful to Neil McNaughton for stimulating discussions of the neurology of the BIS, as well all other RST-related matters. This paper is dedicated to the memory of Professor Blazej Szymura who stimulated the ideas contained in this paper. I am grateful to Dr Andrew Cooper and Dr Arthur Poropat for commenting on an earlier draft of this paper; and I am appreciative of the insightful comments of two anonymous reviewers, as well as the editor, Marco Perugini.

REFERENCES

Allport, G. W. (1927). Concepts of trait and personality. *Psychological Bulletin*, *24*, 284–293.

Andersen, S. B., Moore, R. A., Venables, L., & Corr, P. J. (2009). Electrophysiological correlates of anxious rumination. *International Journal of Psychophysiology*, *71*, 156–169.

Aron, A. R., Durston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. *Journal of Neuroscience*, *27*, 11860–11864.

DOI: 10.1002/per

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, *6*, 115–116.

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

Baars, B. (1997). *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press.

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Bargh, J. A., & Williams, L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, *15*, 1–4.

Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, *11*, 167–203.

Bayer, H. M., & Gilmcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129–141.

Berridge, K. C., & Robinson, T. E. (1995). The mind of an addicted brain: Neural sensitization of wanting versus liking. *Current Directions in Psychological Science*, *4*, 71–76.

Carver, C. S. (2005). Impulse and constraint: Perspectives from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review*, *9*, 312–333.

Carver, C. S., Johnson, S. L., & Joormann, J. (2008). Serotonergic function, two-mode models of self-regulation, and vulnerability to depression: What depression has in common with impulsive aggression. *Psychological Bulletin*, *134*, 912–943.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319–333.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219.

Corr, P. J. (2003). Personality and dual-task processing: Disruption of procedural learning by declarative processing. *Personality and Individual Differences*, *33*, 1–25.

Corr, P. J. (2006). *Understanding biological psychology*. Oxford: Blackwell.

Corr, P. J. (2008). (Ed.), *The reinforcement sensitivity theory of personality*. Cambridge: Cambridge University Press.

Corr, P. J. (2010). The psychoticism-psychopathy continuum: A model of core neuropsychological deficits. *Personality and Individual Differences*, *48*, 695–703.

Corr, P. J., & Matthews, G. (2009). Editors' general introduction. In P. J. Corr, & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. xxii–x1ii). Cambridge: Cambridge University Press.

Corr, P. J., & McNaughton, N. (2008). Reinforcement sensitivity theory and personality. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 155–187). Cambridge: Cambridge University Press.

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*, 603–616.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 304–317.

Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schürhoff, F., et al. (2003). Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: The role of the antérior cingulate. *Proceedings of the National Academy of Sciences USA*, *100*, 13722–13727.

De Wall, F. (2000). *Chimpanzee politics: Power and sex among apes*. Baltimore, MD: The John Hopkins University Press.

Eisenberg, N. (2002). Emotion-related regulation and its relation to quality of social functioning. In W. W. Hartup, & R. A. Weinberg (Eds.), *Child psychology in retrospect and prospect: The Minnesota Symposium on child psychology* (Vol. 32, pp. 133–171). Mahwah, NJ: Lawrence Erlbaum Associates.

Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, *28*, 404–416.

DOI: 10.1002/per

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, *49*, 709–724.
Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.
Farnham, S. D., Greenwald, A. G., & Banaji, M. R. (1999). Implicit self-esteem. In D. Abrams, & M. Hogg (Eds.), *Social identity and social cognition* (pp. 230–248). Oxford: Blackwell.
Farrar, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. *Consciousness and Cognition*, *12*, 609–619.
Feng, W., Luo, W., Liao, Y., Wang, N., Gan, T., & Luo, Y.-J. (2009). Human brain responsivity to different intensities of masked fearful eye whites: An ERP study. *Brain Research*, *1286*, 147–154.
Ferbinteanu, J., & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, *40*, 1227–1239.
Floden, D., & Stuss, D. T. (2006). Inhibitory control is slowed in patients with right superior medial frontal damage. *Journal of Cognitive Neuroscience*, *18*, 1843–1849.
Frith, C. (2007). *Making up the mind: How the brain creates our mental world*. Oxford: Blackwell.
Gazzaniga, M. (1988). *Mind matters: How mind and brain interact to create our conscious lives*. Boston, MA: Houghton Mifflin.
Gomez, R., & Cooper, A. (2008). Reinforcement sensitivity theory and mood induction studies. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 291–316). Cambridge: Cambridge University Press.
Goodale, M. A., & Milner, A. D. (2006). Sight unseen: Two visual systems, one brain. *The Psychologist 19*, 660–663.
Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system* (1st ed.) Oxford: Oxford University Press.
Gray, J. A. (2003). How are qualia coupled to functions. *Trends in Cognitive Science*, *7*, 192–194.
Gray, J. A. (2004). *Consciousness: Creeping up on the hard problem*. Oxford: Oxford University Press.
Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system* (2nd ed.). Oxford: Oxford University Press.
Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.
Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1999). Central serotonin depletion impairs both the acquisition and performance of a symmetrically reinforced go/n-go conditional visual discrimination. *Behavioural Brain Research*, *100*, 99–112.
Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory. *Behavioural Biology*, *12*, 421–444.
Hofmann, W., Gschwendner, T., Friese, M., Wiers, R. W., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*, 962–977.
Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139–154.
James, W. (1884). What is an emotion? *Mind*, *9*, 188–205.
Jermann, F., Van der Linden, M., Adam, S., Ceschi, G., & Perroud, A. (2005). Controlled and automatic uses of memory in depressed patients: Effect of retention interval lengths. *Behaviour Research and Therapy*, *43*, 681–690.
Langens, T. A. (2007). Congruence between implicit and explicit motives and emotional well-being: The moderating role of activity inhibition. *Motivation and Emotion*, *31*, 49–59.
LeDoux, J. E. (1994). Emotion, memory and the brain. *Scientific American*, *270*, 50–59.
Libet, B. (1982). Brain stimulation in the study of neuronal functions for conscious sensory experiences. *Human Neurobiology*, *1*, 235–242.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*, 529–566.
Libet, B. (2003). Timing of conscious experience: Reply to the 2002 commentaries on Libet's findings. *Consciousness and Cognition*, *12*, 321–331.

DOI: 10.1002/per

Libet, B. (2004). *Mind time: The temporal factor in consciousness*. Cambridge, MA: Harvard University Press.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 199–249). San Diego, CA: Academic.

Lykken, D. T. (1995). *The antisocial personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayr, U. (2004). Conflict, consciousness, and control. *Trends in Cognitive Sciences*, *8*, 145–148.

McAdam, D. P., & Pals, J. L. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*, 204–217.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, *96*, 690–702.

McNaughton, N., & Corr, P. J. (2004). A two-dimensional neuropsychology of defense: Fear/anxiety and defensive distance. *Neuroscience and Biobehavioral Reviews*, *28*, 285–305.

McNaughton, N., & Corr, P. J. (2008a). The neuropsychology of fear and anxiety: A foundation for Reinforcement Sensitivity Theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 44–94). Cambridge: Cambridge University Press.

McNaughton, N., & Corr, P. J. (2008b). Animal cognition and human personality. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 95–119). Cambridge: Cambridge University Press.

Metcalfe, J., & Mischel, W. (1999). A hot/cool system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, *106*, 3–19.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, *72*, 1024–1027.

Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, *379*, 449–451.

Morris, R. G. M., & Hagan, J. J. (1983). Hippocampal electrical activity and ballistic movement. In W. Seifert (Ed.), *Neurobiology of the hippocampus* (pp. 321–331). London: Academic Press.

Ortony, A., Norman, D. A., & Revelle, W. (2005). Affect and proto-affect in effective functioning. In J. M. Fellous, & M. A. Arbib (Eds.), *Who needs emotions? The brain meets the machine* (pp. 95–199). New York: Oxford University Press.

Ouimet, A. J., Gawronski, B., & Dozois, D. J. (2009). Cognitive Vulnerability to anxiety: A review and an integrative model. *Clinical Psychology Review*, *29*, 459–470.

Ramachandran, V. S. (2003). *The emerging mind*. London: Profile Books.

Ramachandran, V. S. (2005). *Phantoms in the brain*. London: Harper Collins.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Review*, *18*, 247–291.

Robinson, M. D., & Sedikides, C. (2009). Traits and the self: Toward an integration. In P. J. Corr, & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. 457–472). Cambridge: Cambridge University Press.

Rolls, E. T. (1999). *The brain and emotion*. Oxford: Oxford University Press.

Rothbart, M. K., & Bates, J. E. (2006). Temperament in children's development. In W. Damon, & R. Lerner (Book Eds.) and N. Eisenberg (Vol Ed.), *Handbook of child psychology: Vol. 3, social, emotional, and personality development* (6th ed., pp. 99–166). New York: Wiley.

Rothbart, M. K., Sheese, B. E., & Conradt, E. D. (2009). Childhood temperament. In P. J. Corr, & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. 177–190). Cambridge: Cambridge University Press.

Schell, T. L., Klein, S. B., & Babey, S. H. (1996). Testing a hierarchical model of self-knowledge. *Psychological Science*, *7*, 170–173.

Schneider, W. M., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Schultheiss, O. C., Dargel, A., & Rohde, W. (2003). Implicit motives and sexual motivation and behaviour. *Journal of Research in Personality*, *37*, 224–230.

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

Soubrie, P. (1986). Reconciling the role of central serotonin neurons in human and animal behaviour. behavior. *Behavioral and Brain Sciences*, *9*, 319–325.

DOI: 10.1002/per

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behaviour. *Personality and Social psychology Review*, *8*, 220–247.

Toates, F. (1998). The interaction of cognitive and stimulus-response processes in the control of behaviour. *Neuroscience and Biobehavioral Reviews*, *22*, 59–83.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition*, *15*, 75–118.

Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceeding of the National Academy of Sciences of the United States of America*, *104*, 15011–15036.

Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences*, *14*, 651–726.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*, 181–192.

Wolpe, J., & Lang, P. J. (1977). *Manual for the fear survey schedule*. San Diego, CA: Educational and Industrial Testing Service.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, *27*, 623–633.

Zhu, J. (2003). Reclaiming volition: An alternative interpretation of Libet's experiment. *Journal of Consciousness Studies*, *10*, 61–77.

DOI: 10.1002/per